SCIENCE LIBRARY

JUL 12 1923

# The Journal of Applied Psychology

VOL. VII MARCH, 1923 No. 1

EDITORS

JAMES P. PORTER
Ohio University

WILLIAM F. BOOK
Indiana University

And a Board of Co-operating Editors

PUBLISHED BY THE EDITORS
INDIANA UNIVERSITY PRESS
Bloomington, Indiana

Entered as second class matter November 21, 1922 at the post office at Bloomington, Indiana, under the act of March 3, 1879.
Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917, Authorized December 21, 1922.

# The Journal of Applied Psychology


## CO-OPERATING EDITORS

W. V. BINGHAM, Carnegie Institute of Technology; JUNE E. DOWNEY, University of Wyoming; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston; H. L. HOLLINGWORTH, Columbia University; JOSEPH JASTROW, University of Wisconsin; HERBERT S. LANGFELD, Harvard University; JAMES H. LEUBA, Bryn Mawr College; HENRY C. MCCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, University of Wisconsin; EDWARD K. STRONG, JR., Carnegie Institute of Technology; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; J. E. WALLACE WALLIN, Teachers College, Miami University; G. M. WHIPPLE, University of Michigan; HELEN THOMPSON WOOLLEY, Vocation Bureau, Cincinnati.


The scope of the JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

---


VOLUME VII MARCH, 1923 NUMBER 1


## TABLE OF CONTENTS


A. T. POFFENBERGER. The Conditions of Belief in Advertising ..... 1

H. D. KITSON, C. W. BARNES and C. IUFFENLATZ. Minor Studies in the Psychology of Advertising ..... 10

EVERETT F. PATTEN. An Experiment in Testing Engine Lathe Aptitude ..... 16

JOHN F. MCFADDEN and J. F. DASHIELL. Racial Differences as Measured by the Downey Will-Temperment Test ..... 30

META L. ANDERSON. A Study of the Data on the Results Gathered from Repeated Mental Examinations of 200 Defective Children Attending Special Schools Over a Period of Eight Years ..... 54

G. M. RUCH and M. C. DEL MANZO. The Downey Will-Temperament Group Test; Analysis of Its Reliability and Validity ..... 65

W. T. ROOT. The Freshman: Thorndike College Entrance Tests, First Semester Grades, Binet Tests ..... 77

ARTHUR I. GATES. The Reliability of M.A. and I.Q. Based on Group Tests of General Mental Ability ..... 93

NEWS AND COMMENT ..... 101

BOOK REVIEWS ..... 103


Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this JOURNAL; in every case the year, volume and pages should be specified.


THE JOURNAL OF APPLIED PSYCHOLOGY is published quarterly. The subscription price is $4.00 a year. Single numbers $1.25. Remittances and business communications should be addressed to the Editors, James P. Porter, Ohio University, Athens, Ohio or to William F. Book, Indiana University, Bloomington, Indiana.

# TABLE OF CONTENTS

META ANDERSON
A study of the data on the results gathered from repeated mental examinations of 200 defective children attending special schools over a period of eight years . . . . . 54-64
ADA HART ARLITT
Intelligence tests versus entrance examinations as a means of predicting success in college 330-338
CORYNE BASCH AND HAROLD E. BURTT
Legibility of Bodoni, Baskerville Roman, and Cheltenham Type Faces . . . . . 237-245
MARIAN A. BILLS
Relation of mental alertness test score to positions and permanency in company . . . . . 154-156
HAROLD E. BURTT AND J. CAMDEN CLARK
Facial expression in advertisements . . . . . 114-126
HAROLD E. BURTT AND F. W. IVES
Vocational tests for agricultural engineers . . . 178-187
HAROLD E. BURTT AND CORYNE BASCH
Legibility of Bodoni, Baskerville Roman, and Cheltenham Faces . . . . . 237-245
J. CAMDEN CLARK AND HAROLD E. BURTT
Facial expressions in advertisements . . . . . 114-126
LAURA C. CRAYTOR AND C. B. DAVENPORT
Comparative social traits of various races . . . 127-134
J. F. DASHIELL AND JOHN F. MCFADDEN
Racial differences as measured by the Downey Will-Temperament Test . . . . . 30-53
C. B. DAVENPORT AND LAURA C. CRAYTOR
Comparative social traits of various races . . 127-134
M. C. DEL MANZO AND G. M. RUCH
The Downey Will-Temperament Group Test: a further analysis of its reliability and validity . 65-76
GUSTAVE A. FEINGOLD
The relation between the intelligence and vocational choices of high school pupils . . . . . 143-153
WILLIAM S. FOSTER
Experiments on rod-divining . . . . . 303-311
R. B. FRANKEN AND A. T. POFFENBERGER
A study of the appropriateness of type faces . . 312-329
ARTHUR I. GATES
The unreliability of M. A. and I. Q. based on group tests of general mental ability . . . . . 93-100
B. F. HAUGHT
A scheme for combining incomplete rankings . . 168-172
C. L. HUFFAKER
A contribution to the technique of partial correlation . . . . . 135-142
CLARK L. HULL
Prediction formulae for teams of aptitude tests . 277-284

F. W. IVES AND HAROLD E. BURTT
Vocational tests for agricultural engineers . . . 178-187
GRACE H. KENT
A combination mental test for clinical use . . . 246-257
H. D. KITSON, C. W. BARNES AND C. IUPPENLATZ
Minor studies in the psychology of advertising . 10-15
F. B. KNIGHT AND H. H. REMMERS
Fluctuations in mental production when motivation is the main variable . . . . . 209-223
ARTHUR W. KORNHAUSER
A re-interpretation of the statistical method of Army Trade Tests . . . . . 339-348
DONALD A. LAIRD
The basis of tooth paste sales in representative communities . . . . . 173-177
JOHN H. MCFADDEN AND J. F. DASHIELL
Racial differences as measured by the Downey Will-Temperament Test . . . . . 30-53
BLANCHE M. MINOGUE
A case of secondary mental deficiency with musical talent . . . . . 349-352
ELSIE MURRAY
Freshman tests in the small college . . . . . 258-276
EVERETT F. PATTON
An experiment in testing engine lathe aptitude . 16-29
A. T. POFFENBERGER
The conditions of belief in advertising . . . . 1-9
The return coupon as a measure of advertising efficiency . . . . . 202-208
Measuring the comprehension of advertisements 364-369
A. T. POFFENBERGER AND R. B. FRANKEN
A study of the appropriateness of type faces . . 312-329
H. H. REMMERS AND F. B. KNIGHT
Fluctuations in mental production when motivation is the main variable . . . . . 209-223
W. T. ROOT
The freshman: Thorndike College Entrance Tests, first semester grades, Binet Tests . . . . . 77-92
G. M. RUCH AND M. C. DEL MANZO
The Downey Will-Temperament Group Test; a further analysis of its reliability and validity . . 65-76
IRENE CASE SHERMAN
The Franz Dot Tapping Test as a measure of attention . . . . . 353-359
A. J. SNOW
An advanced course in experimental psychology applied to advertising . . . . . 360-363
Labor turnover and mental alertness test scores . 285-290
M. LAVINIA WARNER
Influence of mental level in the formation of boys' gangs . . . . . 224-236
J. U. YARBROUGH
A study of the correlation of college students' estimates of intelligence with the Otis and other scales . . . . . 157-167

THE
JOURNAL OF APPLIED PSYCHOLOGY
VOL. VII MARCH, 1923 No. 1

# THE CONDITIONS OF BELIEF IN ADVERTISING

A. T. POFFENBERGER
Columbia University

A reader of the literature dealing with the Psychology of Advertising finds numerous articles written upon the effectiveness of various mechanical devices, the attention and memory value of size of space and the position on the page, the influence of color, style of type and its arrangement, the effectiveness of repeating the advertisement, etc. The problem of arousing the confidence of the consumer in the article advertised, the conditions on which it depends, how belief in advertisements may be created and how it may be measured have been very lightly touched in experimental studies. The importance of these matters is emphasized in every textbook on advertising. The fact that the American people are each year induced to squander many millions of dollars in worthless securities through the medium of advertising in some form, and that warnings seem quite ineffective in protecting them, makes one curious about the basis of belief in advertising. It is not enough to say that the American people like to be fooled and that there is no scheme too wild to arouse the confidence of a large proportion of them. The advertiser should know that action is dependent upon belief and that belief in advertising depends upon conditions, some of which at least are under his control.

Belief is indeed a complex mental state and it may depend at any time upon a great variety of factors, the most common of which are listed in textbooks of psychology and advertising. It is not the purpose of this report to repeat such lists but to cite some experimental evidence from the field of advertising to show the importance of a few of the conditions of belief for success in advertising.

1. Belief is rarely the result of reasoning. One does not go through the processes of logic to establish his beliefs. If logic is used at all it is to justify a belief already established. A strik-

ing illustration of the separation of reasoning and belief is found in the case of the insane patient who firmly believed himself to be the son of a king, and yet whose reason was intact enough to enable him to solve complicated mathematical problems.

The advertising of the New Gillette razor offered a good opportunity for studying the relation between belief and reasoning.[1] On May 16, 1921, the Gillette Razor Company announced "a new triumph of American inventive genius of startling interest to every man with a beard to shave." The advertisements state that the "fulcrum shoulder, overhanging cap and channeled guard" were the three innovations which made possible "for the first time in any razor, micrometric control of blade position." A diagram showed "how the blade is biflexed between overhanging cap and fulcrum shoulder. It is flexed once into the inside curve of the cap. This is the minor flexure—the curve for easy gliding action and play of the wrist in shaving. It is flexed a second time—more sharply and in a shorter radius—by the grip of the overhanging cap the whole length of the fulcrum shoulder. This is the major flexure." This arrangement provided an exactness of adjustment to 1-1000 of an inch. Advertisements containing the above information and well illustrated were given to fifty-seven men, college students and university graduates, together with a series of seven questions intended to test both their belief in the new razor and their understanding of it. The answers to these questions showed that all the students agreed that the new razor was better than the old one, and that they would rather pay $5.00 for the new one than $1.00 or $2.00 for the old one. In supporting their belief they were allowed to consult the advertisement as much as they wished. They quoted the "fulcrum shoulder, overhanging cap and channeled guard," which made possible "micrometric control of blade position," but not one of them could explain how the micrometric control was obtained or what advantage there would be in having such micrometric control. They believed that the "channel guard" was an improvement although they could not tell why it was an improvement. As to the importance of major and minor flexures they were entirely ignorant. Five minutes examination of an enlarged diagram of the new razor improved their understanding of the razor little or not at all. Here is a belief effective for the purpose of the advertiser in spite of the inability of the reader to support his belief with reason. This experiment is quoted not to show that the advertisement is poor, but merely to show that in advertising reason-

[1]The study about to be described is part of a larger investigation conducted by David Goldstein and the writer, which will be reported at a later date.

ing is not needed to create belief. Whether the space in the advertisement devoted to argument might have been more profitably filled, is however an interesting question.

2. It is not necessarily that which is true that is believed. The fact that a statement in an advertisement is true will not guarantee belief on the part of the readers of it. The truth may be too startling and surprising to be believed, and in some cases it might be more effective to tell half the truth than the whole truth. Three advertisements were chosen for investigation in this connection, each of them presenting rather unusual information. In each case the truth of the statements made was vouched for by reliable individuals. Each of the advertisements was presented to 100 people, together with a series of statements about them. The individuals were to simply check those statements that most nearly represented their opinion about the advertisements. The statements were carefully prepared so as to avoid suggestion. The studies of the three advertisements will be reported separately:

*Taylor Trunk Advertisement.* This advertisement showed the photograph of a huge elephant standing on a trunk. The picture was accompanied by a signed statement as to the genuineness of the photograph, and by statements to the effect that the trunk was taken from the regular stock. Statistical treatment of the replies of the persons tested showed that 38 percent of them doubted the truth of the statements made in the advertisement; 24 per cent questioned the genuineness of the photograph; and 21 percent believed that it would be impossible to construct a trunk strong enough to withstand such a weight. That is, in from one-fifth to two-fifths of the persons tested the advertisement created a state of mind adverse to the purchase of the article.

*Stanley Vacuum Bottle Advertisement.* This advertisement represented a vacuum bottle falling from a high window, and described an incident in which a Stanley bottle had accidentally fallen from an eighth story window and suffered only a dent, which did not at all interfere with its usefulness. The replies to the questions submitted with this advertisement showed that 31 percent doubted the truth of the statements, while 22 percent refused to believe that a vacuum bottle sufficiently strong to survive the fall could possibly be constructed.

*Edison Phonograph Advertisement.* This advertisement pictured one of the well known "tone tests" demonstrations in Carnegie Hall, Pittsburg. It showed the stage occupied by a well known singer and an Edison phonograph, with the audience in the background. On the advertisement there were reproduced a number of clippings from the daily papers describing the remarkable demonstration, each clipping making the statement in one

form or another that no one in the audience of 2,600 people could distinguish between the voice of the singer and its reproduction. Replies from 100 people to eight questions about this advertisement may be summed up as follows: 77 percent doubted the truth of the statements contained in the advertisements. 73 percent believed that they could tell the difference between a real voice and its reproduction by any phonograph. 82 percent believed that mechanical sounds would betray the phonograph in these "tone tests." 49 percent believed that some form of trickery or deception was practiced. 50 percent believed that the artist intentionally imitated the phonograph. 44 percent believed that a specially constructed phonograph was used for the demonstration rather than a stock instrument. 68 percent believed that *no* phonograph could successfully undergo such a test. 39 percent stated that an actual demonstration of the "tone test" such as that described in the advertisement would not convince them of their inability to distinguish between the human voice and its reproduction.

This advertisement seemed especially effective in failing to create belief in the minds of its readers, when over three-fourths of them doubted the truth of the statements. The writer recommended in this case a less extreme and less startling form of appeal which should not arouse the antagonism manifested in this experiment.

Thus far the negative side of the question has been presented. If belief in an advertisement does not depend upon the truth of the statements made and does not depend upon the reasoning of the reader, on what does it depend? To state the matter simply, we may say that ideas which are present in the mind and are not interfered with by any opposing ideas will be believed. This is merely a bare statement of the law of suggestion and to comply with it in advertising, conflicting ideas should be prevented from entering the mind. There are many conditions on which such undisturbed acceptance of ideas depends. Only three will be mentioned here.

1. The ideas aroused by an advertisement must not conflict too sharply with the reader's experience. Introspections volunteered by many of the 100 subjects who served in the experiments just described indicated that their past experiences with trunks, vacuum bottles and phonographs furnished conflicting ideas which the advertisements were not sufficiently powerful to overcome. This was especially true in the case of the phonograph advertisement where doubt was expressed in a large percentage of the cases.

An experimental study of an advertising campaign that failed

showed clearly the need for complying with the conditions of belief. "Radior Products," a series of toilet preparations were introduced into this country by an English firm. The appeal contained in the advertising may be illustrated by the following quotation taken from one of the advertisements: "Radior is the magic new word in the book of beauty. It means the triumphant union of the finest complexion preparations with actual radium. Its content of radium works the miracle of nature. It purifies the skin, gives it the health to regain its youthfulness and loveliness." The container of these preparations was represented as emanating rays which were very suggestive of bolts of lightning. These toilet preparations could not be sold—the campaign had failed. Why? One possible reason for the failure was that the suggestion to apply radium on the skin for toilet purposes conflicted too sharply with people's opinions about radium. In order to discover what people believed about radium, a questionaire was carefully prepared in the form of a True-False test and submitted to 400 people, comprising a well-to-do, well-educated group, a group of average intelligence and financial standing and a special group of persons working in "Beauty Parlors." The questionaire is reproduced below. It was constructed so as to minimize the influence of suggestion and to correct for it when it did occur, as for instance in combining replies from question 2 and 7. Questions 3, 5 and 10 were introduced as fillers and had no special significance.

Read the first statement on this sheet. If you *feel* that it is true, underline the word "True." If you feel that it is false, underline the word "False." Then judge each of the other statements as true or false. If you do not know, guess at it. This sample statement is false and is marked as it should be:

There are eight days in the week.....True False

1—Radium is a deadly poison.....True False

2—Radium is so costly that it cannot be used for commercial purposes.....True False

3—Treatments with radium will make the hair grow .....True False

4—Radium is used in the treatment of such diseases as cancer.....True False

5—Radium is a substance which gives off light.....True False

6—Radium when applied in diluted form gives a healthful stimulation to the skin.....True False

7—Radium is now used in the manufacture of many products in common use.....True False

8—Radium causes burns when it comes into contact with the skin.....True False

9—Toilet preparations containing radium are to be recommended for their beneficial effects.....True False
10—Certain forms of radium can be obtained at little cost .....True False
11—Radium preparations should not be used except on the advice of a physician.....True False
12—Radium products are gaining popularity as mild stimulants to health and growth.....True False

Read the following question and answer it by underlining "Yes" or "No":

If you should be told that there was a toilet preparation of high quality and moderate price and which contained a minute quantity of radium, would your knowledge of the effects of radium be such as to make you want to buy it?.....Yes No

Nothing need be said here concerning the difference in the reactions of the different groups to the questionaire, except that the most intelligent group had the strongest reaction against radium-containing products, and that the "Beauty Parlor" workers had the least antagonism. Inquiry among these people indicated that some of them had confused radium with the violet ray which was at the time a very popular form of treatment with them. The results of the study of the questionaire may be summarized as follows:

Forty-eight per cent of the persons tested believed that radium is a deadly poison; 80 per cent believed that it causes burns when it comes in contact with the skin; 89 per cent believed that radium preparations should be used only upon the advice of a physician; 71 per cent said that they would not buy any kind of toilet preparation that they knew contained even a minute quantity of radium; 90 per cent thought of radium as a substance used for the treatment of cancer. The whole experiment may be summed up in the statement that about 70 per cent of all the replies indicated opinions unfavorable to radium-containing products as toilet preparations. The basis of this unfavorable reaction could easily be traced to the people's accumulated experiences of radium. Clippings of all articles dealing with radium and appearing in newspapers and magazines were collected for a short period. Practically all of them emphasized the harmful effects of radium and the dangers incurred in handling it, instead of any beneficial properties it might possess.

This is a clear case of the inability of an advertising campaign to overcome the resistance established by experience. To create a favorable attitude toward radium-containing products as toilet preparations by an educational campaign conducted in news-

papers and magazines might conceivably have been possible. But it certainly would have been impracticable.

The mass of experiences upon which beliefs rest is gradually accumulating and being modified by additions. As some one has said, "Confidence is at the same time stable and fragile." It is stable because of the bulk of experiences on which it rests; it is fragile because it is subject to change from new experiences. A supposed bargain which is found to be not a bargain, a derogatory newspaper report about a store, a misprint in the advertised price of an article; any one of these may be the experience that will introduce a conflicting idea into the mind of a consumer and upset the conditions favorable for his belief. The advertiser, appealing as he does to populations, cannot take account of the shiftings of individual experience and belief, but he must take account of the widespread and slowly accumulating experiences which may amount to powerful prejudices against his appeal.

2. Ideas that are to create belief must come from an authoritative source. This is a well-known law of suggestion. The hypnotist can do nothing without his air of authority and the subject's recognition of it. We are accustomed to believe the statements made by a person in whom we have confidence, and to believe what is printed in a medium which we consider authoritative. Even if there is conflict with one's own experience, he will sometimes accept the contrary experience of another person as a basis for belief if he have sufficient confidence in the other person. But even then the new experience may not be too conflicting. Advertisers have for years striven to develop an atmosphere of confidence and authority by all the devices at their command. The present experiment was intended to measure in a tentative fashion the degree of confidence which an advertisement can create in comparison with other forms of publication. I have compared the degree of belief or doubt aroused by the three advertisements previously described (namely Taylor Trunks, Stanley Vacuum Bottles and Edison Phonographs) with the belief aroused by essentially the same statements coming from a reputable journal. In order to make such a comparison, the facts stated in each advertisement were prepared in the form of a news item abstracted from an engineering magazine. One such abstract will illustrate the character of all. These abstracts were presented to a group of 100 persons of the same general character as those tested in the earlier experiments, but who knew nothing of those experiments or nothing of the purpose of the present experiment. Along with the abstract of each advertisement was a series of questions as nearly as possible like those used in the test with the advertisements. As far as we were able to ascertain, no one doubted the

authenticity of the news abstracts. Instead of reporting the results of the three questionaires in detail, it will be sufficient to compare for each of the three cases the number of replies indicating doubt in the advertisement and doubt in the news item. It must be remembered that the news item mentioned no trade-named article nor indicated any connection with specially advertised goods.

The following is an extract from————————Magazine of June, 1921. Read it carefully and then answer the questions given below, by scratching out with a pencil the part of each that does *not* give your opinion.

A Remarkable Shock Resisting Construction

"A new type of wardrobe trunk was recently put through a remarkable test of strength. The five-ton elephant belonging to the New York Hippodrome was made to stand on the empty trunk. An examination of the trunk following the test showed that it had withstood the great shock without damage.

Interesting photographs were made of the test, showing the elephant mounting the trunk, seated upon it, and then stepping off of it."

1—I believe—doubt—do not believe—the statements made in the above extract.

2—I feel that the photographs that portray the test are—are not—genuine.

3—I do—do not—believe that a trunk could be made strong enough to successfully undergo the test described.

Percentage of replies expressing

| doubt | *In the Ads* | *In the News Items* |
|---|---|---|
| Taylor Trunks | 28 per cent | 43 per cent |
| Stanley Vacuum Bottles | 27 per cent | 43 per cent |
| Edison Phonographs | 55 per cent | 40 per cent |

In two of the cases it will be noted that there was greater confidence in the advertisement than in the news clipping, while in the third there was greater confidence in the news clipping. An examination of the three advertisements did not afford an entirely satisfactory explanation for the shift of belief in the case of the third advertisement. It seems safe to conclude from these records that although belief in certain advertisements may be low, they may carry at least as much authority in presenting a set of facts as can be conveyed by a news article. The doubt in the cases we have studied is the effect rather of conflict of ideas with experience, than the effect of the use of an unauthoritative medium of expression. Introspections volunteered by the subjects suggest that

illustrations and especially photographs used in the advertisements tend to strengthen belief. Unless one suspects trickery, as some of our subjects did, looking at a picture ought to carry with it greater weight in establishing belief than merely reading printed matter. Even if Mark Twain was right in advising that one believe only half that he sees and nothing that he hears, the advantage in favor of the picture is obvious.

3. There is a third important condition of belief, namely, that we tend to believe what arouses our desires, our fears and our emotions generally. I have no experimental evidence to offer in this connection, and know of none in the field of advertising. But evidence for the importance of this factor may be drawn from psychology. Wm. James has said, "A man who has no belief in ghosts by daylight will temporarily believe in them when alone at midnight, he feels his blood curdle at a mysterious sound or vision, his heart thumping and his legs impelled to flee." In strong emotion we might find the condition responsible for the belief in the bargain counter. The politician finds no difficulty in honestly believing what best fits in with his aspirations, while his opponent may as honestly believe the opposite and for a like reason. If one really wants a certain suit of clothes or an automobile which costs more than he should pay, he may honestly believe that he is making an economical purchase. Once a belief has been established in this way, logic and reasoning may be used to support it.

These three conditions of belief which I have described represent three important variables in the formula for advertising success. For most advertising situations they are unknown variables which may interact in a very complex manner. But they may be discovered by careful examination. That they are not always sought or discovered, is clear from the cases I have cited which were taken from advertising already used. Three questions might well be asked about every advertisement: (1) What adverse beliefs and experiences does it have to meet in the minds of consumers? (2) Will the authority which it wears by its mode of presentation or by the medium in which it appears enable it to create belief? (3) Does the appeal used arouse desires which will in turn create belief in the advertised article? These are human behavior questions that psychological methods will enable the advertiser to answer before the advertising is used as well as after the money has been spent upon broadcasting it.

# MINOR STUDIES IN THE PSYCHOLOGY OF ADVERTISING

From the Psychological Laboratory of Indiana University
By H. D. Kitson, C. W. Barnes, and C. Iuppenlatz

## IX. Right and Left Hand Pages in Magazines

A considerable amount of discussion has centered around the question as to which is the more favorable position for advertisements in magazines, the right or left-hand page. The question has usually been approached through the laboratory method, subjects viewing the momentarily exposed pages of a "dummy" or a real magagine and then writing down all they have seen in the order in which they have seen it. With this technic Adams[1] reports a value of the right hand page which is to that of the left as 100 is to 65. That is, he estimates the right hand to be fifty per cent superior. Hotchkiss and Franken[2] estimate from their investigation that the right-hand page of a magazine is worth five per cent more than the left-hand page.

In order to secure additional light upon the question the writers approached the problem with the historical technic, seeking to discover the frequency with which advertisers actually used the right-hand and left-hand pages.

The materials used were 1082 consecutive pages of advertising matter in the American Magazine and 1000 in the Saturday Evening Post, for the year 1922. Beginning with the first advertisement following solid reading matter tabulation was made of the successive pages containing advertising matter, exclusive of cover pages. Results are shown in Table I.

TABLE I.

SHOWING THE FREQUENCY WITH WHICH RIGHT AND LEFT-HAND PAGES CARRY ADVERTISEMENTS IN THE AMERICAN MAGAZINE AND THE SATURDAY EVENING POST FOR 1922

| | Total Right | | Total Left | | Total | Full Right | | Full Left | | Full Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | | No. | % | No. | % | No. |
| Amer. ..... | 576 | 56 | 506 | 44 | 1082 | 301 | 82 | 64 | 18 | 365 |
| S. E. P..... | 521 | 52 | 481 | 48 | 1000 | 350 | 67 | 171 | 33 | 521 |

[1] H. F. Adams, Advertising and its Mental Laws, New York, Macmillan, 1916. p. 100.
[2] G. B. Hotchkiss and R. B. Franken, Attention Value of Advertisements, New York, New York University, 1920, p. 24.

These figures show that in both magazines the right-hand page is used for advertisements more frequently than the left; in the American 56 per cent of the times and in the Saturday Evening Post 52 per cent of the times; with a frequency on the average 17 per cent greater than that for the left-hand page.

Still stronger evidence of the favor in which the right-hand page is held comes from facts shown in a separate tabulation which was made concerning the location of full page advertisements. There were 365 of these in the American Magazine and 521 in the Saturday Evening Post. Eighty-two per cent of the former and 67 per cent of the latter appeared on the right-hand page. In the first named periodical the right-hand page was used for full page advertisements five times as frequently as the left, and in the Saturday Evening Post twice as frequently. It was observed that certain advertisers who consistently used the same relative position in issue after issue almost invariably used the right-hand page.

On the assumption of the historical method, that a practice much in vogue has probably proved itself profitable, it would appear from these results that the right-hand page in magazines is more profitable than the left. True, the practice here revealed need not necessarily be considered as having been based upon conscious reckoning of comparative profits and losses from right and left-hand insertions. Even at best it probably represents only the fruits of a "sub-conscious" "hunch." Nevertheless it is coming to be recognized that such a process of summing up experience has some validity as a guide in business practice where scientifically valid measures are not yet obtainable.

Appearance on the right or left-hand page of a magazine should, however, not be considered as wholly a result of the voluntary choice of the advertiser. Requirements of journalistic make-up would inevitably put some advertisements on the right and some on the left-hand page. Again, the requirements of color process or of some other technical matter might force certain advertisements on one page rather than another. Moreover, some advertisers do not specify either page, not regarding the matter as important enough to bother with. For these reasons the more frequent use of the right-hand page can not be regarded as purely a reflection of profitable experience. Still these factors alone might be expected to work toward an equal distribution of advertisements upon the right and left-hand pages. The fact that they are unequally distributed indicates that some selective factor is at work in favor of the right-hand page. And the fact that this page is used with overwhelming frequency by the purchasers of the expensive full page advertisements indicates that the right-hand

page has demonstrated its superiority to at least some influential and successful advertisers. On slighter evidence than this space-sellers have set "preferred-position" prices.

## X. Position on Newspaper Page

In answer to the question, what is the best portion of a newspaper page for an advertisement, recommendations have been made as follows: the upper half, the right-hand portion, upper left-hand corner, upper right-hand corner. Or, in recognition of the need for considering the right and left-hand pages separately, the statement has been made that the upper outside corner is preferable.[1] These conclusions have been reached chiefly from laboratory investigations in which the subjects have been shown a "dummy" or a real newspaper or magazine containing advertisements or nonsense syllables distributed among the various portions of each page, and have been asked later to write down those items which they have remembered. The positions of those items mentioned the greatest numbers of times have been judged to be the most favorable. It should be noted that the conclusions regarding favorability of positions are for the most part based upon studies made with magazines. There is considerable doubt whether conclusions reached regarding positions in magazines are true also for newspapers.

Regardless of the advantageousness of these particular positions on a newspaper page it may be seen from even a cursory examination of the best type of newspaper that they are not equally open to the use of advertising matter. For there is a practice of newspaper make-up that crowds advertisements into certain portions of the page in order to reserve certain other portions for news. A circumstance which presents grave obstacles to the use of the positions which have been so laboriously designated as advantageous.

Specifically described, the practice of newspaper make-up, technically called "pyramiding," is to block up the advertisements first from the lower right-hand corner of the page. On some pages (usually the second page of the newspaper) only a small portion of this corner is used. But as an increasingly larger portion of the page (towards the middle and latter half of the paper) is consigned to advertisements the advertisements begin to run upward and leftward until often only a small portion of the upper left-hand corner is left for news. But this corner is usually thus

[1]Adams, op. cit. p. 101.
Hotchkiss and Franken, op. cit. pp. 23-25.
D. Starch. Advertising. Scott Foresman and Co. 1914, Chicago, pp. 108-115.

reserved (unless the entire page is devoted to advertisements). Here is an apparent block to effective newspaper advertising, at least to that theoretically recommended by laboratory investigators.

In order to measure this block and express it in quantitative terms the writers undertook an investigation of actual journalistic practice. Choosing a newspaper that admittedly represents a high standard of journalistic make-up and advertising value, they tabulated the number of times the different quarters of pages were occupied by advertisements, and from this calculated the chances that an advertiser might secure a "favored" place for his advertisement.

The New York Times was used, and 1000 successive pages bearing advertisements and news were chosen from the issue of May, June and July, 1922. Those pages were omitted which contained nothing but advertisements, also those containing classified advertisements. Notation was made in each case as to whether or not the particular portion (half or quarter) contained any advertising matter, be it ever so small. Results are shown in Table I, and Figure 1.

TABLE I.

SHOWING THE NUMBER AND PER CENT OF UPPER AND LOWER, RIGHT AND LEFT PORTIONS OF LEFT AND RIGHT HAND PAGES OCCUPIED BY ADVERTISEMENTS IN THE NEW YORK TIMES DURING MAY, JUNE AND JULY, 1922.

| 505 Left Hand Pages | | | | 495 Right Hand Pages | | | |
|---|---|---|---|---|---|---|---|
| Lower Half | | Upper Half | | Lower Half | | Upper Half | |
| Left Quarter | Right Quarter | Left Quarter | Right Quarter | Left Quarter | Right Quarter | Left Quarter | Right Quarter |
| No. % | No. % | No. % | No. % | No. % | No. % | % No. | % No. |
| 175 35 | 390 77 | 150 30 | 370 73 | 115 23 | 335 68 | 115 23 | 295 60 |

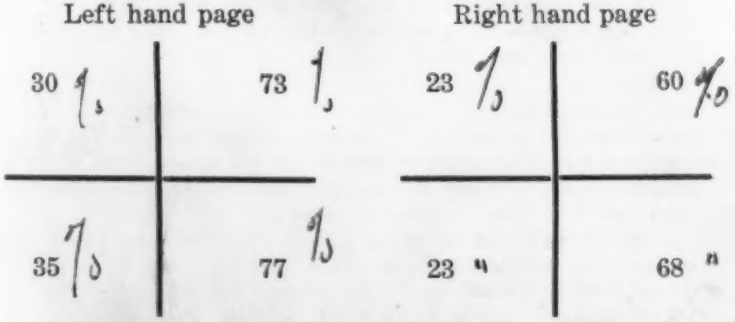



Fig. 1. Showing the per cent of pages on which the various quarters were occupied by advertising matter.

*1. Left and Right-hand Pages.*

Of the one thousand pages examined, 505 were on the left-hand page and 495 on the right-hand page. (It is interesting to note in this connection that *full-page* advertisements, which had been tabulated incidentally but not counted among the thousand shown in the table, appeared more frequently on the right-hand page. Of a total of 235 full-page advertisements, 120 or 53 per cent were on the right-hand page.)

*2. Lower half.*

*a. Left-hand page.* On the 505 left-hand pages the lower left quarter was occupied by advertisements in 175 or 35 per cent of the cases; the lower right quarter, in 390 or 77 per cent of the cases.

*b. Right-hand page.* On the 495 right-hand pages the lower left quarter was occupied by advertisements in 115 or 23 per cent of the cases; the lower right quarter, in 335 or 68 per cent of the cases.

*3. Upper half.*

*a. Left-hand page.* On the 505 left-hand pages the upper left quarter was occupied by advertisements in 115 or 23 per cent of the cases; the upper right quarter, in 295 or 60 per cent of the cases.

*b. Right-hand page.* On the 495 right-hand pages the upper left quarter was occupied by advertisements in 155 or 23 per cent of the cases; the upper right quarter, in 295 or 60 per cent of the cases.

## CONCLUSIONS.

These results show that regardless of the alleged superiority of certain positions in newspapers, there are pronounced limitations to the likelihood that these positions may be secured. Thus although the right-hand page may be more desirable, it will be open for advertisers no more frequently than the left, since these results show it to be occupied only 495 times in 1000.

The upper outside quarter, which is recommended as being most valuable of all can be secured only 30 times in 100 on the left-hand page, though, fortunately it can be secured 60 times in 100 on the right-hand page. (For the sake of easier visualization, the frequency with which the various quarters of left and right-hand pages are used is shown diagramatically in Figure 1.)

The right-hand side of both pages is twice and even three times as likely to carry advertisements as the left-hand side. Which means that in the case of the left-hand page the majority of advertisements are going to be on the inside. And, unfortunately for the advertisers, this tendency is more marked in the case of the left-hand page than in the case of the right. These amounts

may not hold good for all newspapers, especially for papers using the pyramid system more thoroughly than the New York Times. With such papers it is likely that the percentage of pages carrying advertisements in the upper left-hand corner would be still smaller.

From a journalistic point of view this quantitative investigation of the practice of pyramiding shows that the pyramiding progresses much more rapidly upward than leftward, the upper right portion of the page being occupied almost as frequently as the lower right portion. Though it should be remarked that the advertisements did not usually occupy the entire upper right quarter; tabulation was made if an advertisement merely extended into the area.

While this investigation furnishes no light concerning the relative values of the various positions on the newspaper page, it shows that there are distinct limitations to the ease with which certain positions can be secured, and it shows, at least for one high-grade newspaper, the size of the limitations in quantitative terms. It also suggests that the findings of laboratory studies regarding many other problems of advertising might profitably be checked up with actual conditions in order to show the practicability of their application in concrete situations.

# AN EXPERIMENT IN TESTING ENGINE LATHE APTITUDE[1]

By EVERETT F. PATTEN
University of Wisconsin.

The investigation reported in this paper was carried out in co-operation with the Shop Practice Department of the College of Engineering, University of Wisconsin. It was able to profit materially by a preliminary investigation made along the same lines in 1921.[2] It was also favored by the fact that in the Shop Practice Course there were many students turning out supposedly identical pieces of work on fairly similar lathes. There was an added advantage in the fact that the men directing the work in the shop were university trained, and consequently were able to give such aid that the preliminary psychological analysis of the job was made more readily than would probably be the case in industry.

The particular type of work investigated was a course in elementary lathe practice. This involved four hours of shop work every week and two hours of instruction every two weeks. The students, following a combination of oral and printed instructions together with blueprints, turned out five pieces on an engine lathe during the semester. These were: a worm, a worm-wheel blank, a bevel-gear blank, a spiral-gear blank, and a planer bolt. The first four were turned out of cast iron and the last from rod steel.

The criterion of this investigation was the amount of skill developed in one semester of instruction. More specifically, the criterion was the quality of lathe work produced by the students during the course. Instructors' estimates are notoriously unreliable, as was found by the preliminary investigation and should never be used if an objective criterion can be obtained. Fortunately in this case the pieces themselves could be measured and a thoroughly objective criterion thus found in the accuracy with which these pieces correspond to the blueprint specifications. Working with great care Mr. Brecky of the Shop Practice Department took thirty-three micrometer measurements on the five pieces produced by each man measuring the various important dimensions

[1]From the Psychological Laboratory, University of Wisconsin. Grateful acknowledgment is hereby made to Prof. Clark L. Hull, Department of Psychology, and to Prof. W. L. Dabney, director of Machine Practice. Prof. Hull has supervised the investigation and given much helpful criticism, especially in the preparation of the manuscript. Prof. Dabney has given hearty co-operation and been of great assistance particularly in procuring subjects and in perfecting some of the tests.

[2]This investigation was reported in a master's thesis by D. W. Nelson. It is on file in the library, University of Wisconsin.

of these pieces. Several thousand very accurate readings were thus obtained. In addition the general finish of each piece was scored on a percentage basis.

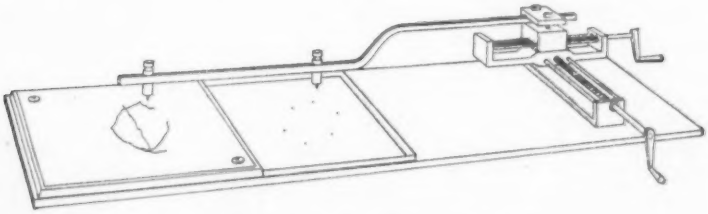
The measurements obtained were reduced to percentage grades. The deviation of each measurement from the dimension called for in the blueprint was first found. A check distribution was then made for each of the thirty-three measurements each distribution containing the deviations of all the men on that measurement. The median of each distribution was assigned the value of 85, the average grade given by the Shop Practice Department. The intervals of the distribution between the median and zero deviations were assigned values between 85 and 100 according to the size of the deviation. The fourth check from the opposite end of the distribution was assigned the value of 70 which is hardly passing on the school scale. The intervals of the distribution between this point and the median were assigned values between 70 and 85. The last three checks of the distribution were given below 70 or failure. Thus the larger the deviation from the standard the smaller would be the percentage grade. The mean of the grades for each individual piece was found and the average of the mean grades of the five pieces belonging to one man was computed in order to get his final grade in lathe working ability. Although this process took weeks of work the resulting criterion was very reliable. Indeed the large coefficient of total correlation (R) obtained later was probably due in no small degree to the care taken in obtaining the criterion.

In proceeding to find tests which would measure lathe ability it was first found necessary to analyze that ability as far as possible. This was done by watching the men at work observing particularly wherein the good ones differed from the poor ones. As a result of this analysis eight tests were selected. Some of these at first glance would seem to have little connection with lathe ability in general, but they were included because they were thought to test a constituent factor. In selecting the tests an especial effort was made to get tests as dissimilar as possible so that there would be little duplication. The detailed analysis will be given with the discussion of the various tests.

(A) While studying the men at work it was noticed that the worker must be able to estimate two dimension space and to co-ordinate the motions of his two hands. He must also be quick and sure of movement. In order to test these abilities the "co-ordination machine" was devised.[3] The apparatus is shown in Fig.

[3]The need for such a machine was first seen by Dr. Hull, who worked out its general principles. The practical details of construction were worked out by Prof. Dabney. He also designed and made a number of improved models.

1. It may be best thought of as a compound lathe "rest" attached



Coordination Machine. Fig. I.

to a board. The action of the machine depended on two screws placed at right angles to each other. Turning these screws by means of small cranks caused a carriage to move in any desired direction in a horizontal plane. A long tracer arm was fastened firmly to the carriage. Projecting down from the middle of the arm was a small brass point. This touched the surface of a hard rubber plate which was fastened to the board near the main part of the machine. Embedded in the plate so that their tops were flush with the surface were six brass pegs arranged in a three-inch hexagon. These pegs were connected by an electric wire so that when the tracer point touched one of them a buzzer sounded. The object of the test was to start on signal from a certain point and make the circuit of the six brass spots as quickly as possible. The time taken to go around was measured by a stop-watch. A correlation of this with the criterion gave a-.423. This was the largest and most valuable correlation obtained.

(B) Another measure was taken of this test, that of distance traveled. The tracer arm extended out beyond the contact point and had a pencil fastened to its extremity. This traced an exact record of the circular path on a piece of paper inserted for the purpose. A typical record is shown in Fig. 2. The distance traversed by this rather meandering line was measured by a specially devised instrument. The frame supporting the works of an ordinary watch was filed so as to expose the gear wheel on the shaft of the small second hand. The hands of the watch remained intact. The frame was so held in the hand of the measurer that the teeth of this gear could be run around the pencil tracing, following in detail all its turns. The distance traversed by this wheel was shown by the hands of the watch and could be read off at a glance. It will be seen that the circular distance was measured in units of revolutions of the small hand. One minute (or one revolution) was equal in length to 29.5 millimeters. The correlation of this measure with the criterion was +.192. This was rather sur-

prising since it meant that the greater the distance, the better the grade on shop work. However, it was included in the final team of tests.

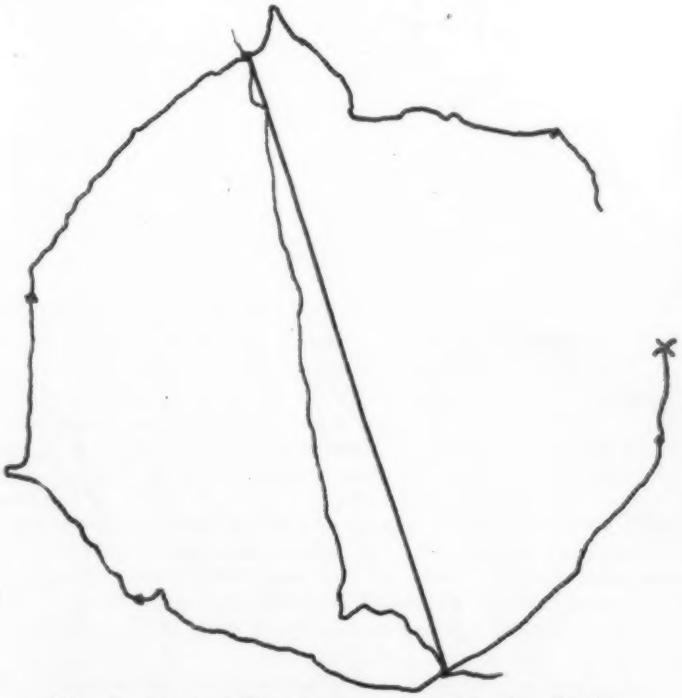


Fig. 2. Typical Tracing from Coordination Machine.

(C) A third score was also secured by the use of the co-ordination apparatus described above. It was observed that the worker must frequently cut a true surface in a diagonal direction using his eye to guide him. Accordingly the subject was required to run the tracer diagonally across between two of the most distant pegs going as quickly as possible. Here, as before, the time was taken with a stop-watch. The correlation of this measure with the criterion gave –.356. This was large but since it was thought that it would duplicate the first measure taken, that of time on the circular distance, measure C was discarded.

(D) The second measure of the diagonal co-ordination was made in deviations from a straight line between the two pegs. If a subject traced a straight line his record would coincide with

the diagonal and there would be no deviation. However, if he wavered and made wrong moves, his record would deviate considerably from the straight line distance and would be poor in proportion to the amount of deviation. This deviation was measured by the area of surface included between the record tracing and the true diagonal as shown by a straight line drawn between the points. This was measured by means of a planimeter. The correlation of this measure and the criterion was –.034. This was so small as to be worthless. If the instructions had been to trace as straight a line as possible this might easily have proven a valuable test. It seems to have distinct possibilities.

It will be noted that the co-ordination test described above is a form of miniature test. It attempted as is usual in such cases to incorporate in a single test as many of the various elements necessary for success on the job and in as nearly the right proportions, as possible. Being a miniature it was portable and easily handled. One defect of the test is that it is undoubtedly influenced by training and thus is not a pure aptitude test.

(E & F) Further analysis of the job showed the necessity for close co-ordination between hand and eye as well as for quickness and accuracy of movement. Accordingly the number checking test was introduced. This was the standard Woodworth and Wells number checking test in which the subject was started on signal and told to cross out all groups of digits containing both 4 and 7. He was given one minute. The correlation of the number of right groups checked with the criterion was +.055. This was too low to be of use so was discarded from the final team of tests. The number of groups wrongly checked correlated –.035 so it was also discarded.

(G) It was believed that the possession of mechanical ingenuity would be important for success in lathe working so an effort was made to test for this quality. Accordingly an adaptation of Stenquist's lock assembly test was used. The lock was simple and inexpensive. It consisted of six parts which were: a rectangular lower case or plate, a dog, a spring, a bolt, a screw and an upper plate. The subject was shown a similar lock which was already assembled and told to put the parts of the other lock together like it. Although the lock seemed very simple, many had great difficulty in assembling it. Four failed entirely giving up after working from eleven to twenty-one minutes. The correlation with the criterion was –.199. Although not very high (–.33 was obtained last year) it had such a low correlation with the other tests that it was kept in the final team as tapping a real constituent of the particular aptitude under consideration.

(H) The ability to follow printed directions and blueprints is obviously a factor in learning lathe work. To measure this ability and to gain knowledge of the subject's carefulness and painstaking traits, a box-directions test was introduced. This test is the well-known Healy puzzle box, consisting of a small black cabinet containing a lock mechanism worked by three levers at the top and left sides of the box. The object of the test is to open this door. The necessary steps in the process were printed on a large card bearing an excellent diagram of the box. This was before the subject as he worked. This test proved even more difficult than the lock test. Ten men did not open the box at all, some giving up with the first failure. The correlation of the time with the criterion was –.270. This was large enough to warrant its use in the final team of tests.

(I) The number of wrong moves made was another measure taken on this test. However, this measure undoubtedly did not introduce any new element because errors caused the subject to start over again and thus increased his time. In addition the correlation was low (–.112) so this measure was eliminated.

(J) In doing good lathe work a great deal depends on the skill of the operator in taking accurate measurements. Accordingly the cylinder gauging test was introduced. A series of nine brass cylinders was very carefully made under Prof. Dabney's direction. These measured .970″, .965″, .960″, .940″, .930″, .900″, .885″, and .870″. It will be seen that there were three groups each composed of three cylinders. Each group or series was graduated by a different constant the same constant being used throughout a single group. The finest series changed by steps of .005″, the next by .010″, and the coarsest by .015″. The subject was given a six-inch machinist's rule, and a pair of calipers and told to find the cylinders which measure .965″ and .885″ as quickly as possible. The time taken to find the two cylinders constituted one measure. This correlated –.127 with the criterion and was too low to be of value.

(K) Another score on the cylinder test was the amount the chosen cylinder deviated in measurement from the one asked for. Thus if the subject chose the .930″ cylinder in place of the .965″ he was given a deviate score of .035. If he thought the .900″ cylinder measured .885″, he scored .015. The sum of these errors gave his deviation score, as .050. This measure correlated –.196 with the criterion. This seemed sufficiently high to be of use but it was finally decided to eliminate the test from the final team because it was really a trades test measuring acquired skill rather than aptitude.

(L) An effort was made to test ingenuity and reasoning ability by the match-squares test, though at the outset there seemed but a gambling chance of success. The match-squares test consisted of seventeen matches glued to a board in such a way as to make six squares in two banks of three each. The subject was shown these squares then given a mimeographed sheet on which short lines were drawn to represent the matches in the same position as on the board. He was told to cross out five of the short lines in such a way that if they were removed there would be left exactly three whole squares and no lines left over. The time was taken with a stop-watch. When through, the subject was asked whether he had ever done this or a similar puzzle before. The fact that many were already familiar with it together with the additional fact that twenty-two out of the sixty subjects could not do it at all seems to have destroyed what little diagnostic value the test may have had. The correlation with the criterion was only +.164. This was too small and also in the direction contrary to what would have been expected so the test was discarded.

(M) In watching the men at work it seemed probable that a keen sense of space relationship was necessary, particularly in guiding the cutting tool to its place and in finding centers. Accordingly the circle centering test was devised. A sheet of paper bearing a number of mimeographed circles, (see Fig. 3) was put before the subject. He was given a pencil and asked to make a dot in the center of each circle as quickly and accurately as possible. The task was complicated somewhat by the overlapping of some of the circles. The time required to place these dots was taken by a stop-watch and constituted one measure. This correlated –.201 with the criterion and since it was quickly and easily taken. Since the test was so cheap and portable it was kept in the final team of tests.

(N) Another measure was taken from this test the errors in locating the centers. These deviations from the true centers were measured translated into grades by a rather complicated process and these grades averaged to get the final grade for each man. Since this measure was not used the process will not be discussed further. The correlation with the criterion was fairly high (–.273) but was in a direction contrary to what was expected. Apparently the good subjects sacrificed accuracy to speed. Partly because of this, and partly because of the labor in scoring, this test was not included in the final team.

(O) It seemed reasonable to believe that if a man were skillful in handling woodworking tools he would also tend to be an adept in using other kinds, such as are employed in lathe work. Fol-

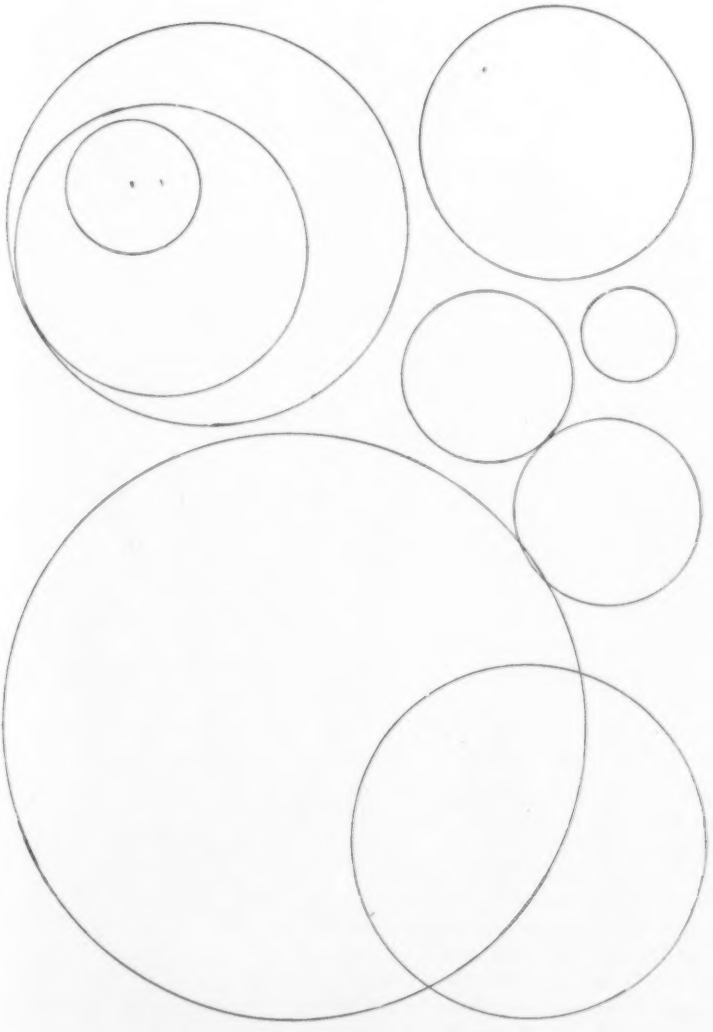


Fig. 3. Circle Centering Test.

lowing this line of reasoning the woodworking test[4] was given. The subject was asked to saw a three-inch square from a small board of definite size and shape, being given a saw, a square, a pencil and a special device for holding the wood such as is in common use in manual training shops. He was timed by means of a stop-watch. This measure had the surprisingly low correlation of –.119 with the criterion. It was discarded.

(P & Q) Two other measures were taken from the woodworking test one in deviation in dimensions from the three-inch square and the other in deviations from the true right angles. These latter measures were very difficult to make, requiring hours of almost microscopic work with a vernier rule, a square and a rule calibrated to one hundredths inch. The correlations in both cases were low, being –.095 and +.076 respectively. Both measures were discarded. It was a source of some surprise that the woodworking test did not have more diagnostic value.

The instructions for each test were given according to an unvarying formula so that the results would be as comparable as possible. The tests were given in a quiet room in a definite sequence. When the subject was admitted he went in regular succession from one test to the next, showing a card containing his name and number to each tester. By such devices a constant stream of subjects was kept flowing smoothly through the testing room. In order to obtain useful results, it was thought desirable to use a minimum of sixty records. Consequently over seventy men were tested so as to allow for the loss of ten through defective records from various causes.

A summary table of the correlations already given is as follows:

TABLE I

| | r |
|---|---|
| *(A) Criterion with co-ordination test, time | –.423 |
| *(B) Criterion with co-ordination test, distance | +.192 |
| (C) Criterion with co-ordination test, diagonal time | –.356 |
| (D) Criterion with co-ordination test, diagonal deviation | –.034 |
| (E) Criterion with number checking, rights | +.055 |
| (F) Criterion with number checkig, wrongs | –.035 |
| *(G) Criterion with lock test, time | –.199 |
| *(H) Criterion with box test, time | –.270 |
| (I) Criterion with box test, errors | –.112 |
| (J) Criterion with cylinder gauging, time | –.127 |
| (K) Criterion with cylinder gauging, deviations | –.196 |

[4]The woodworking test was devised by Mr. F. C. Mueller of the Psychology Department, University of Wisconsin.

TABLE I—cont.

| | |
|---|---|
| (L) Criterion with match square, time | +.164 |
| *(M) Criterion with circle centering, time | –.201 |
| (N) Criterion with circle centering, grades | –.273 |
| (O) Criterion with woodworking, time | –.119 |
| (P) Criterion with woodworking, deviations in dimensions | –.095 |
| (Q) Criterion with woodworking, deviations in angles | +.076 |

The variables finally used were renumbered for convenience:

| | |
|---|---|
| Criterion | =1 |
| Co-ordination, distance | =2 |
| Co-ordination, time | =3 |
| Lock test, time | =4 |
| Box test, time | =5 |
| Circle centering, time | =6 |

The means and standard deviations of these variables are shown in Table II.

TABLE II.

| *Means.* | *Standard Deviations.* | *Unit of Measurement.* |
|---|---|---|
| $M_1$=84.046 | $S.D._1$=3.689 | Percentage grade on work pieces |
| $M_2$= 6.9406 | $S.D._2$= .2653 | Minutes-distance** |
| $M_3$= .9256 | $S.D._3$= .3869 | Minutes |
| $M_4$= 5.8585 | $S.D._4$=5.089 | Minutes |
| $M_5$= 4.78 | $S.D._5$=2.627 | Minutes |
| $M_6$= .3765 | $S.D._6$= .2547 | Minutes |

The zero order correlations and intercorrelations are shown in Table III.

TABLE III.

$r_{12}$=+.192

$r_{13}$= –.423 $r_{23}$= –.229

$r_{14}$= –.199 $r_{24}$=+.142 $r_{34}$=+.056

$r_{15}$= –.270 $r_{25}$=+.118 $r_{35}$=+.036 $r_{45}$=+.103

$r_{16}$= –.201 $r_{26}$=+.022 $r_{36}$=+.170 $r_{46}$= –.032 $r_{56}$=+.059

The P. E. of each of the correlations in the first column is roughly .08.

It will be noted that although the correlations with the criterion are fairly high, the intercorrelations are exceptionally low. More than usual stress was laid on the latter in this investigation because it was thought that this principle is generally observed too little. To get the undiluted correlations with the influence of the other tests removed, the zero order correlations were subjected to

*Indicates the tests which were included in the final team. The above r's were all product-moment coefficients.

**Distance in terms of circumference of gears on second hand.

partial correlation, using Yule's formula[4] and Rosanow's[5] general technique.

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

There were some fifty-two of these partials. Only those of highest order are given in Table IV:

TABLE IV

Partial correlations

$r_{12.3456} = +.194$

$r_{13.2456} = -.363$

$r_{14.2356} = -.211$

$r_{15.2346} = -.285$

$r_{16.2345} = -.100$

These partial correlations show but very little shrinkage and show that the experiment was unusually successful securing a set of tests highly uncorrelated with each other.

After the partial correlations were found, the coefficient of total correlation was computed. This correlation of the criterion with all the tests weighted in the best possible manner, was computed in two ways to check the arithmetic.[6] The formulae were

$$R = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13.2}^2)(1 - r_{14.23}^2)(1 - r_{15.234}^2)(1 - r_{16.2345}^2)}$$

$$R = \sqrt{1 - (1 - r_{16}^2)(1 - r_{15.6}^2)(1 - r_{14.56}^2)(1 - r_{13.456}^2)(1 - r_{12.3456}^2)}$$

One correlation was .5549 and the other .5545 showing all the partials to be correct. This correlation is about as large as seems possible to secure from such highly selected groups as university students. It demonstrates the possibility of detecting aptitude in quality of lathe work produced.

In order to get the maximum yield from a team of tests as well as to use a set of test scores to predict probable success in a given vocation, a regression equation must be obtained. The

[4]Yule, Introduction to the Theory of Statistics. p. 239.
[5]Rosanow, Analysis of Mental Function. p. 39-43.
[6]Rosanow, op. cit., p. 40.

general form[1] for this equation for six variables (criterion and five tests) is:

$$x_1 = r_{12.3456}\frac{SD_{1.23456}}{SD_{2.13456}}x_2 + r_{13.2456}\frac{SD_{1.23456}}{SD_{3.12456}}x_3$$

$$+ r_{14.2356}\frac{SD_{1.23456}}{SD_{4.12356}}x_4 + r_{15.2346}\frac{SD_{1.23456}}{SD_{5.12346}}x_5$$

$$+ r_{16.2345}\frac{SD_{1.23456}}{SD_{6.12345}}x_6$$

Where $x_1$= the deviation of any subject from the mean of that variable (here the criterion), where $x_2$, $x_3$, $x_4$, $x_5$, and $x_6$ equal the deviation of any subject from the means of the respective variables, and where $S.D._{1.23456}$, etc., equal the partial standard deviations. These partial standard deviations are obtained by a formula of the type:[2]

$$S.D._{1.23456} = S.D._1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}\sqrt{1-r^2_{15.234}}\sqrt{1-r^2_{16.2345}}$$

The partial standard deviations are shown in Tabie V:

TABLE V.

| | |
|---|---|
| $S.D._{1.23456}$=3.070 | $S.D._{4.12356}$=4.881 |
| $S.D._{2.13456}$= .248 | $S.D._{5.12346}$=2.483 |
| $S.D._{3.12456}$= .343 | $S.D._{6.12345}$= .247 |

Substituting these, together with the partial correlation coefficient, in the formula for the regression equation and solving we have:

$x_1$=2.397$x_2$–3.252$x_3$–.132$x_4$–.353$x_5$–1.996$x_6$

This, however, is not a very convenient form of the regression equation for practical use. Substitutions were accordingly made

[1] Yule, op. cit., p. 244.
[2] Yule, op. cit., p. 237.

for the small x's so that instead of deviations from means were had individual scores minus the mean of that variable ($X_n$-$M_n$). The formula then becomes:

$$X_1-M_1=2.397(X_2-M_2)-3.252(X_3-M_3)-.132(X_4-M_4)-.353(X_5-M_5)$$
$$-1.996(X_6-M_6)$$

Substituting values for the various means (found in Table III) and solving, the following equation was secured:

$$X_1=73.635+2.397X_2-3.252X_3-.132X_4-.353X_5-1.997X_6$$

It will be noted that the values $X_2$ to $X_6$ represent gross scores of any given subject on the present team of tests while $X_1$ represents his probable grade in a course such as previously described. Now if a person of unknown aptitude in this kind of work should be given the above tests, his test scores might be substituted in the above equation and his most probable grade in such a course be predicted with some precision. To take an actual case, subject number 9 made in this team of tests, the following scores:

Co-ordination, distance=6.6 revolutions
Co-ordination, time=1.05 minutes
Lock test=1.17 minutes
Box test=3.37 minutes
Circle centering, time=.35 minutes

Substituting these in the equation, the result is:

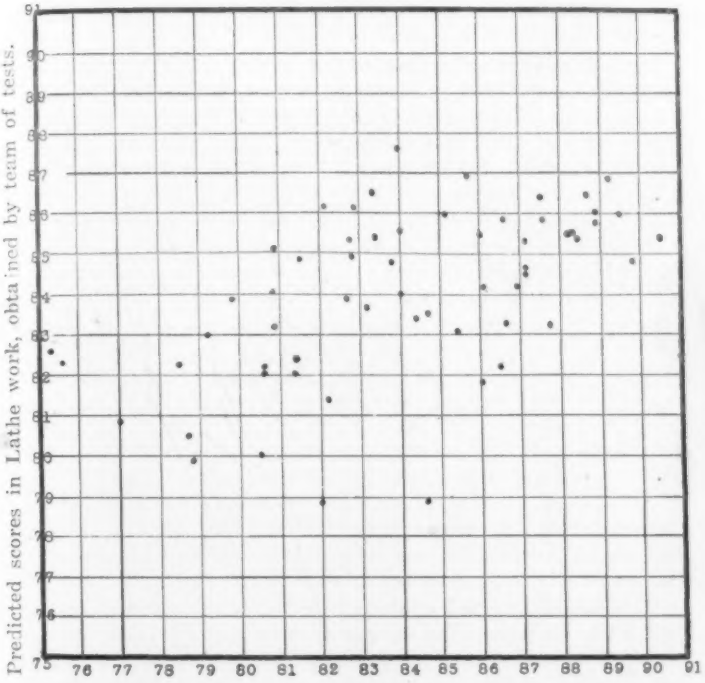
$$X_1=73.635=2.397(6.6)-3.252(1.05)-.132(1.17)-1.996(.35)$$

Solving, $X_1=84.008$. This man actually scored 84 in his five pieces.

To lighten the work of prediction, both for this investigation and for the one to be carried on with a new group of subjects here next semester, a chart was prepared containing all the products of the probable test scores and their respective weights. The scores were then predicted for all the sixty subjects, both to confirm the R and to observe the detailed relation of individual predicted scores to the actual scores. This is shown in detail in the scatter diagram, Fig. 4.

FIGURE 4.

Scatter Diagram.



# RACIAL DIFFERENCES AS MEASURED BY THE DOWNEY WILL-TEMPERAMENT TEST

By JOHN H. MCFADDEN and J. F. DASHIELL

## INTRODUCTION

Study of character and personality traits is, by its very nature, a perplexing problem, since there are a multiplicity of factors entering into the make-up of any given individual, and since the organization of these factors varies so greatly with each individual. Several tentative classifications of the traits and of the factors governing the combinations of the traits have been made. Woodworth[1] analyzes personality into four factors: Physique, Chemique or Temperament, Instincts, and Intelligence. F. H. and G. W. Allport[2] divide tendencies into: Intelligence; Temperament, including Emotional Breadth and Emotional Strength; Self-expansion, including Extro-Introversion, Ascendance-Submission, Expansion-Reclusion, Compensation, and Insight and Self-Evaluation; and Sociality, including Social Participation, Self-seeking and Aggressive Self-seeking, and Susceptibility to Social Stimuli. Tests for these were given, and the results checked by estimates of character and personality given by friends of the subjects tested. The above division of tendencies is considered by the writers as merely a tentative one which may be of some use in guiding further character analysis. Watson[3] gives a form for analysis of personality, and Wells[4] classifies the healthy mental reactions as distinguished from unhealthy ones. H. T. Moore and A. R. Gilliland[5] studied aggressiveness as a personality trait, isolating it as far as possible by selecting two groups of students, one composed of students rated as the most aggressive in their class, and the other composed of students rated as the least aggressive in their class. The following tests are suggested as measurements of aggressiveness: eye movements; distraction from staring, from electrical shock, and from visual presentation of a snake; and a

[1] Woodworth, R. S. *Psychology, a Study of Mental Life.* Henry Holt and Co., New York. 1921.

[2] Allport, F. H. and Allport, G. W. *Personality Traits: Their Classification and Measurement.* Journal of Abnormal Psychology and Social Psychology. Vol. 16, No. 3. April, 1921.

[3] Watson, J. B. *Psychology from the Standpoint of a Behaviorist.* J. B. Lippincott Co., Philadelphia.

[4] Wells, F. L. *The Systematic Observation of the Personality—In Its Relation to the Hygiene of the Mind.* Psychological Review, Vol. 21, 1914, pp 295-332.

[5] Moore, H. T. and Gilliland, A. R. *The Measurement of Aggressiveness.* Journal of Applied Psychology. Vol. 3, No. 2, June, 1921.

word association test. In the above tests, the first group scored high, and the second group scored less, thus indicating the value of the methods used. R. O. Filter[6] studied Self-Assurance and Speed of Decision, concluding that one test was not sufficient for each trait, but that two or more should be given. Also, he concludes that there was not a very close correlation between speed of decision and self-assurance. The test used in the present study of racial differences is the Downey Will-Temperament Test[7] formulated by Dr. J. E. Downey, of the University of Wyoming.

The particular problem here studied is the comparison of temperament as found in individuals of the white and negro races. Previous studies have been made regarding the comparative intelligence of the two races. Mayo[8] finds that, while the "data points clearly to a measurable degree of mental difference," the differences are of degree, and not of kind, and that mental variability is slightly less in negro high school students than in white high school students. Ferguson[9], using the Woodworth and Wells Mixed Relations Tests I and II, a form of the Ebbinghaus Completion Test, a Cancellation Test, and one of the Columbia Maze Tests, finds the intellectual performance of the general colored population to be approximately seventy-five per cent as efficient as that of whites, with the possibility of negroes more nearly approaching whites in reasoning based upon concrete objects than in reasoning based upon mental representations of the objects. He finds a practical equality in sense capacity in the perceptive and discriminative abilities, and, since the difference must exist in the central elaborative powers upon which thought more directly depends, he infers that such neural differences as may be found will probably be mainly in the constitution of the cortical neurons, rather than elsewhere in the nervous system. This point will be taken up later. Ferguson also states that, while no certain conclusion as to racial variability can be reached, it would be reasonable to suppose that the negroes would vary more than whites, since the so-called negro race is not nearly as homogeneous as the

[6] Filter, R. O. *An Experimental Study of Character Traits.* Journal of Applied Psychology, Vol. 5, No. 4, December, 1921.

[7] Downey, J. E. *The Will-Profile, A Tentative Scale for Measurement of the Volitional Pattern.* Univ. of Wyoming Bulletin, Vol. 15, No. 6A, Feb., 1919.

*The Adolescent Will-Profile.* Journal of Educational Psychology. Vol. 11, No. 3, March, 1920.

*Some Volitional Patterns Revealed by the Will-Profile.* Journal of Experimental Psychology, Vol. 3, No. 4, August, 1920.

*Manual of Directions for Testing.* World Book Co., Yonkers-on-Hudson, N. Y., 1921.

[8] Mayo, M. J. *The Mental Capacity of the Negro.* Archives of of Psychology. Monograph No. 28, 1913.

[9] Ferguson, G. O., Jr. *The Psychology of the Negro.* Archives of Psychology, Monograph No. 36, April, 1916.

white race. Derrick[10] finds that the negro college student is older than the white, that their average intelligence quotient is nine points lower than the average white intelligence quotient, and that the white college student is about seventy-seven per cent as variable as the negro college student.

Turning from previous work on intelligence, and considering other studies of racial differences, we find several reports on different traits. Stetson[11] finds practically no difference in memory in the two races, though there may be some doubt as to the value of this contribution. Bache[12] finds the negro quicker in his reaction time than the white man, and Garth[13] finds that there are slight differences in their work curves—that there is a tendency of the negros to do more of their work in the first half of the allotted time than for the white workers. This would indicate that the negroes are more quickly aroused to work, but that, after the first spurt, their effort diminishes. Mitchell, Rosanoff and Rosanoff[14] in a study of association, find that negro children, on the whole, show further departure than white ones from the normal adult association standard.

Before going on into our particular problem, it is necessary to note some cautions which experimental work has given. Pressey[15] warns that grade norms may vary very greatly in various schools, that there may be different age distribution, and that other factors may enter in. And, in a comparison of students of different races, it is all the more important to bear this in mind, since grade norms will inevitably vary much more greatly in schools for different races than in different schools for students of the same race. Arlitt[16] finds a tendency for greater variation (in results from intelligence tests) in the same race but different social status than in the same social status but in different races. This latter point will be considered in the next section. Further, as Gatewood[17] points out, there is a tendency to rate negro children above their mental status, as compared with the tendency to rate whites at or below their true mental status.

---

[10] Derrick, S. M. *A Study of the Intelligence of White and Colored College Students.* Journal of Applied Psychology. Vol. 4, pp. 316-329, December, 1920.

[11] Stetson, C. R. *Some Memory Tests of Whites and Blacks.* Psychological Review, Vol. 4, pp. 283-289, 1897.

[12] Bache, R. M. *Reaction Times With Reference to Race.* Psychological Review, Vol. 2, pp. 475-486, 1895.

[13] Garth, T. R. *White, Indian and Negro Work Curves.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[14] Mitchell, A., Rosanoff, I. R., and Rosanoff, A. J. *A Study of Association in Negro Children.* Psychological Review, Vol. 26, September, 1919.

[15] Pressey, S. L. *The Problem of the Unselected Group in the Standardization of Tests.* Journal of Applied Psychology. Vol. 5, No. 1, March, 1921.

[16] Arlitt, A. H. *On the Need for Caution in Establishing Race Norms.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[17] Gatewood, E. L. *Teachers' Estimates of Negroes and Whites.* School and Society, Vol. 9, pp. 90-91, 1919.

## SUBJECTS AND TESTS

The subjects tested were white and negro high school and college students. It was hoped, by choosing such groups, to eliminate many of the factors of selection, and to make the groups more directly comparable, and also to fix a general level of intelligence between the groups. It was found, however, that, in general, negro students of each grade were slightly older than white students of the same grade. This is constantly found in comparisons of white and negro students. Graphs and tables follow, which endeavor to analyze the general results, and to interpret them in terms of equal ages and equal grades. These results are based upon the records of 38 high school and 39 college students from each race.

The thirty-eight white high school students were tested at the Chapel Hill High School, and were chosen at random, nearly all of the male students being taken. A note should be made here concerning the make-up of this high school, since it is a little out of the ordinary. The high school students are drawn from three fairly distinct social classes: those from the surrounding farm and mill districts, those whose parents are in business in the town of Chapel Hill, and those whose parents are members of the Faculty of the University of North Carolina. A survey of general intelligence of these students with regard to their respective social classes was made in 1917 by Chase and Carpenter and again in 1921 by Dashiell and Glenn, who found that the social levels and intelligence levels were strongly related. As might be expected, the Faculty children led in general intelligence, followed by the Town children, with the Out-of-Town group last. Chapel Hill High School embraces the 8th, 9th, 10th, and 11th grades, with seven years of grammar school work preparatory.

The thirty-nine white college students were taken from classes in psychology at the University, the entire class in Abnormal Psychology being tested, and the balance volunteering from introductory classes. Except for the very few freshmen and seniors, it is a fairly typical group of college students.

For the negro subjects, students in three negro institutions were tested. The National Training School, at Durham, N. C., furnished five college students and all except two of the high school students. There is some doubt as to the validity of all the results from this group, since the testing was stopped, after forty subjects had been secured, upon evidence that some of the students were being coached by their fellows. This would naturally affect the Reaction to Contradiction test, and it seemed to affect the Motor Inhibition and Co-ordination of Impulses tests, since there

was an unnatural tendency for the subjects to write excessively slowly in the preliminary warming-up (Test II, 3), and, for the Co-ordination of Impulses, to take great pains in adhering to the space limitations, but to fail in the time limitations. These tendencies were not noted with other groups, either of negroes or of whites. The doubt concerning coaching, and the fact that these tests and the norms for grading them were developed primarily for adults, would render the high school comparisons somewhat inconclusive, and would emphasize the results from the college groups.

The second group of negro students was from Shaw University, in Raleigh, N. C. The men tested were selected by the President of that University. Shaw University is one of the best negro institutions in the South in regard to college ranking. Eighteen men were tested there, during a period of three days. The writer does not believe that any of the men in this group were coached.

The last group of negro students was from the Negro Agricultural and Technical School, at Greensboro, N. C. This is a state institution, but, owing to the fact that agricultural schools usually have lower standards of admission than the Colleges of Liberal Arts in the various Universities, it is possible that a comparison of the men in this group with men from the University would be unfair. Still, the three negro schools visited are typical of the institutions of that race, and the men afford a better basis of comparison, being more widely distributed, than if all the subjects were taken from one institution. At the Agricultural and Technical School, eighteen men were tested during a period of three days, sixteen of the men being of college grade, and two of high school grade.

Although precautions were taken to standardize the subjects in regard to education, it must be admitted that the criticism of Arlitt[18] regarding differences in social status holds good here. Derrick[19] reports that "White boys from average environment go to college nearly twice as frequently as negro boys from average negro environment"; thus, the negro college students are more highly selected from their environment than the white students from theirs. The group of negro agricultural college students furnished a partial check on this condition, but for the best study in race comparison, it would be necessary to select a large number of individuals from all environments. The present comparison can claim to hold good only for college and high school students of the two races.

[18] Arlitt, A. H. *On the Need for Caution in Establishing Race Norms.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[19] Derrick, S. M. *A Study of the Intelligence of White and Colored College Students.* Journal of Applied Psychology, Vol. 4, pp. 316-329, December, 1920.

To better insure uniformity in results, all of the tests here reported were given by one of the writers and were graded by him. The grades were checked by the same writer and by a student trained in the method of grading.

It is not thought necessary to give an exhaustive description of the Will-Temperament Test, but definitions of the traits scored, and a statement of the value of the scores in each, are necessary for a full understanding of the results. In the order in which they are listed by Prof. Downey[20], they are:

*Speed of Movement:* Speed of movement of person relative to size and age. A score of "10" here indicates a very high speed, grading down to "1", which indicates very slow movement.

*Freedom from Load or Inertia:* Tendency to work at one's highest speed without external pressure, little tendency to relax speed, quickness in warming up to a task. A score of "10" indicates practically uniform speed, "1" indicates great increase, which may mean that the subject usually works more slowly than he could—that he has excessive "load" or weight of inertia.

*Flexibility:* Ease and success in readjustment; capacity to modify one's routine reactions. A score of "10" here indicates highest success, "1," practically failure.

*Speed of Decision:* Quickness in reaching a decision or conclusion. A slow reaction here may be due to caution or conservatism in weighing the elements involved in a situation, or be caused by one's being side-tracked by irrelevant matters or a rambling procedure. Here "10" indicates a speedy decision, "1", a very slow decision.

*Motor Impulsion:* This trait refers both to impetuosity and energy of reaction. Consider the ease with which the brakes or inhibitions can be removed, and also the tendency to explosive reaction when the brakes are actually off. A "10" here indicates explosive or expansive reaction under distraction, grading down to a "1," which indicates inhibition or withdrawal into one's self.

*Reaction to Contradiction:* This refers to the degree of confidence with which one maintains his opinion against contradiction. A "9" or "10" reaction signifies an aggressive subject—the burden of proof is thrown on the person who does the contradicting; "7" and "8" are confident reactions, but reasons are cited for one's confidence and the burden of proof is accepted; "5" and "6" are tactful reactions; below "5" there is a grading down to complete failure to assert one's own opinion.

*Resistance to Opposition:* The vigor with which one reacts immediately to a blocking of one's purpose. It grades from "10", a

[20] Downey, J. E. *Manual of Directions for Testing.* World Book Co., Yonkers-on-Hudson, N. Y., 1921.

strenuous reaction, to "1", complete passivity in the face of opposition.

*Finality of Judgment:* Confidence in previous decision, willingness or unwillingness to reopen a question once closed. A "10" indicates refusal to reconsider, or a very rapid reconsideration, a "1" indicates extremely slow reconsideration. Sometimes the subject spends as much time on review as he does on the original decision. The score in this trait should be interpreted while bearing in mind the score or time spent on the original decision.

*Motor Inhibition:* Capacity to keep in mind a set purpose and achieve it *slowly*. It involves power of motor control, imperturbability, and patience. A score "10" indicates highest success, a "1" indicates practically no inhibition.

*Interest in Detail:* Attention to details. This trait is not equivalent to accuracy, but rather it indicates thoroughness, which may or may not be accompanied by analytical thought or insight. A "10" indicates highest care for details, a "1" shows very little care.

*Co-ordination of Impulses:* Capacity to execute a double task without preliminary trial, capacity to handle a complex situation successfully without forgetting either factor involved. This trait is probably allied to keeping one's head in a confusing situation. A "10" here indicates greatest success, a "1" indicates practically failure.

*Volitional Perseveration:* Absorption in a task; willingness to keep "plugging away" at it because the examinee sets up a goal for himself. A "10" here is also the most desirable score, while a "1" indicates little desire to take pains in perfecting a piece of work that the individual takes personal pride in.

The scores in the Downey tests were obtained by taking the records of a large number of individuals, for some given trait, and assigning score values in such a way that 10 per cent of the persons tested attained each score. Thus, in a given group, we could hardly expect to find normal distribution curves for the various traits, but rather a tendency for an equal number to attain each score. Further, the total score made on the test is not as significant as the grouping of the scores in the various traits, although it has been considered that a high total score indicates a "strong" personality, and a low total score indicates a "weak" personality. Also, the tests and grading norms were devised for adults, and this should be taken into consideration when these tests are given to younger subjects.

G. M. Rush[21] gives a criticism of the Will-Temperament

[21] Rush, G. M. *A Preliminary Study of the Correlations Between Estimate of Volitional Traits and the Results from the Downey "Will-Profile."* Journal of Applied Psychology, Vol. 5, No. 2, June, 1921.

Test, in regard to correlation of the results and estimates. He finds, for fifteen subjects, very little correlation between the estimates of the traits and the scores in those traits as tested by the Will-Temperament Test. Prof. Downey[22], however, using a slightly different method of estimates, found sufficient correlation to justify the tests.

## RESULTS AND INTERPRETATIONS

Table I contains the summary of means and medians of the scores in the various traits for the two races, divided into college, high school, and combined groups, with the standard deviation from the mean in the various traits for the two college groups.

TABLE I.

| | College Students | | | High School | | Combined | |
|---|---|---|---|---|---|---|---|
| WHITE STUDENTS | | | | | | | |
| Trait | Mean | S. D. | Median | Mean | Median | Mean | Median |
| Speed of Movement | 6.77 | 2.531 | 7 | 4.13 | 3 | 5.46 | 5 |
| Freedom from Load | 5.41 | 2.527 | 5 | 6.71 | 7.5 | 6.06 | 6 |
| Flexibility | 5.15 | 1.740 | 5 | 4.13 | 4 | 4.65 | 5 |
| Speed of Decision | 4.10 | 2.888 | 4 | 3.21 | 3 | 3.66 | 3 |
| Motor Impulsion | 4.51 | 1.781 | 5 | 3.76 | 4 | 4.14 | 4 |
| Reaction to Contradiction | 5.79 | 2.451 | 7 | 5.50 | 6 | 5.65 | 6 |
| Resistance to Opposition | 5.77 | 2.731 | 7 | 5.50 | 7 | 5.63 | 7 |
| Finality of Judgment | 4.92 | 2.392 | 5 | 5.60 | 5 | 5.25 | 5 |
| Motor Inhibition | 5.41 | 3.409 | 6 | 4.60 | 5 | 5.01 | 5 |
| Interest in Detail | 7.33 | 1.489 | 7 | 6.44 | 6 | 6.89 | 7 |
| Co-ordination of Impulses | 6.02 | 2.882 | 7 | 3.05 | 2.5 | 4.82 | 4 |
| Volitional Perseveration | 6.64 | 2.956 | 8 | 5.81 | 6 | 6.23 | 7 |
| NEGRO STUDENTS | | | | | | | |
| Trait | Mean | S. D. | Median | Mean | Median | Mean | Median |
| Speed of Movement | 4.10 | 3.070 | 3 | 3.08 | 2 | 3.60 | 3 |
| Freedom from Load | 5.48 | 2.716 | 5 | 5.87 | 5 | 5.67 | 5 |
| Flexibility | 4.00 | 2.000 | 4 | 4.23 | 4 | 4.11 | 4 |
| Speed of Decision | 3.87 | 3.048 | 2 | 4.71 | 4 | 4.28 | 3 |
| Motor Impulsion | 5.15 | 1.955 | 5 | 5.47 | 5 | 5.31 | 5 |
| Reaction to Contradiction | 6.10 | 1.997 | 7 | 6.81 | 7 | 6.45 | 7 |
| Resistance to Opposition | 5.33 | 2.641 | 5 | 5.05 | 5 | 5.19 | 5 |
| Finality of Judgment | 4.77 | 2.301 | 5 | 6.29 | 7 | 5.52 | 5 |
| Motor Inhibition | 3.84 | 3.519 | 3 | 6.76 | 8 | 5.28 | 6 |
| Interest in Detail | 6.82 | 2.146 | 7 | 6.52 | 7 | 6.67 | 7 |
| Co-ordination of Impulses | 3.87 | 2.838 | 3 | 2.13 | 1 | 3.01 | 2 |
| Volitional Perseveration | 5.69 | 2.524 | 6 | 5.80 | 7 | 5.74 | 7 |

[22] Downey, J. E. *The Will-Profile, A Tentative Scale for Measurement of the Volitional Pattern.* Univ. of Wyoming Bulletin. Vol. 15, No. 6A, Feb. 1919.

Attention must be called to the fact that a composite profile plotted from mean or median scores, as given in Figure 1 and

FIGURE 1. Mean Scores.

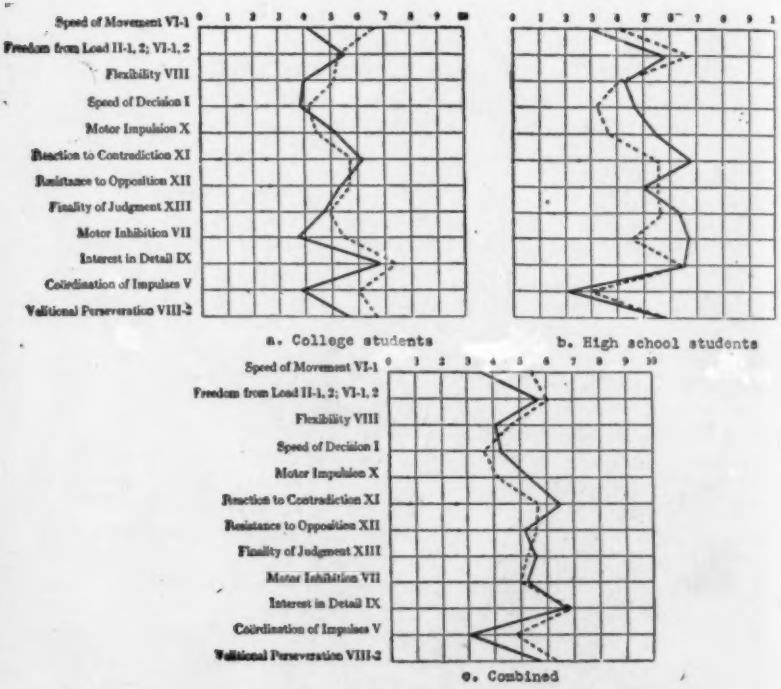
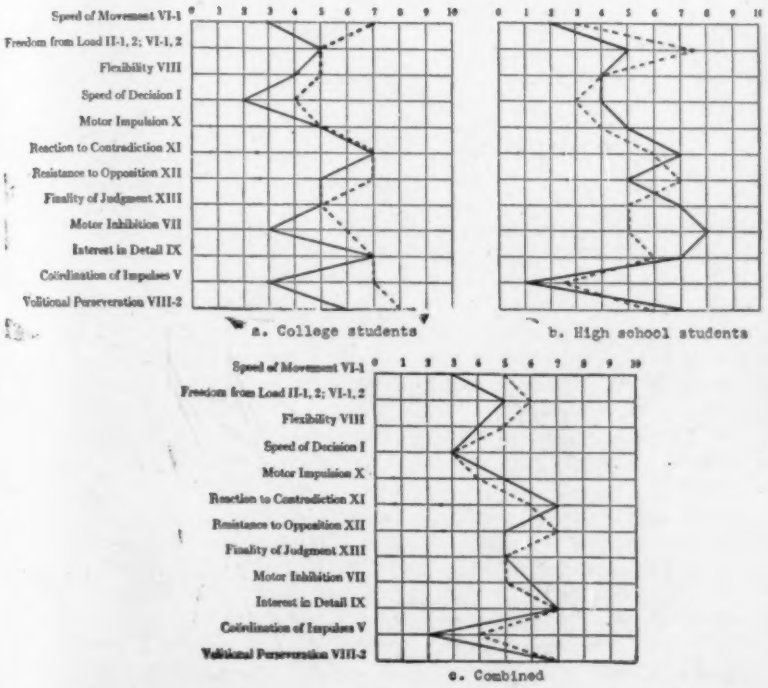

Speed of Movement VI-1
Freedom from Load II-1, 2; VI-1, 2
Flexibility VIII
Speed of Decision I
Motor Impulsion X
Reaction to Contradiction XI
Resistance to Opposition XII
Finality of Judgment XIII
Motor Inhibition VII
Interest in Detail IX
Coördination of Impulses V
Volitional Perseveration VIII-2
0 1 2 3 4 5 6 7 8 9 10
a. College students
b. High school students
c. Combined

Figure 2, must be read with caution, since there is so great possibility of individual variability in the scores in each trait. Also, these profiles do not indicate, more than in a very general way, tendencies toward racial patterns. However, from Table I, and Figures 1 and 2, we may say in general that the "average" white student has greater speed of movement, that he has slightly greater freedom from load—that is, he tends to work on the weighing of the evidence, or less decisiveness; that he is slightly average more nearly at his greatest speed; that he has slightly greater flexibility and ability to meet new conditions in life; that he is slightly slower in making decisions—which may mean a better less liable to become explosive under distraction from conflicting stimuli; that he is less forceful when contradicted, which may mean

a greater degree of social adjustment (a score of "5" is the tactful response), or a greater degree of suggestibility; that he is firmer when confronted by obstacles or physical opposition; that he takes about the same time, or slightly longer in reconsidering

FIGURE 2. Median Scores.


Speed of Movement VI-1
Freedom from Load II-1, 2; VI-1, 2
Flexibility VIII
Speed of Decision I
Motor Impulsion X
Reaction to Contradiction XI
Resistance to Opposition XII
Finality of Judgment XIII
Motor Inhibition VII
Interest in Detail IX
Coördination of Impulses V
Volitional Perseveration VIII-2
0 1 2 3 4 5 6 7 8 9 10
a. College students
b. High school students
c. Combined

his decisions, which may be due to greater care, or to greater indecision; that he has, as a college student, greater ability or desire to restrain his impulses, and, as a high school student, less ability or desire (but here must be remembered the criticism given above, that there was some evidence of coaching upon the results from negro high school students); that he has practically the same interest in detail; that he has a markedly greater ability to coordinate his impulses, or to meet confusing situation; that he has a slightly greater tendency to volitional perseveration.

Figure 3 shows the distribution of the individual scores in each trait, for the high school students, the college students, and the

combined groups, of both races. Score values are indicated on the horizontal axis, number of individuals on the vertical.

The median profile of the high school students (Figure 2b) shows emphasis on the traits which combine to suggest an aggressive personality, a result which was found by Dr. Downey[23] in testing high school freshmen.

A point may here be brought out in relation to the negro's tendency to resist contradiction more strongly than the white. Dr. Bevis[24], tabulating the frequencies of various forms of psychoses among the Southern negroes, reports that dementia praecox stands at the head of the list of their psychoses, with the catatonic form occurring about twice as often as in the white. Since one of the characteristic symptoms of the catatonic form of dementia preacox is negativism, the result of this test checks fairly with Dr. Bevis' report. Also, (and cf. No. 6 in "Summary") we might expect a greater degree of aggressiveness from the negro when he is allowed to reveal his inner self. As a race, they are generally oppressed or held down, a fact which might lead us to look for a reaction in the direction to greater aggressiveness when unchecked. Observations of Northern negroes have been reported, where the negro, freed from race restraints, has swung to the other extreme and has been rather too aggressive; and the tendency of the negro drivers to beat their mules more mercilessly than do white drivers, is a matter of observation. This is apparently a case of "inferiority complex" and "compensation".

The lower score of the white students in motor impulsion might have been expected. Odum[25] says, "The two general characteristics, then, the negro's emotional nature are the lack of restraint and the consequent extreme expression of the feelings." These would naturally be allied with greater motor impulsion.

Various investigators have made different reports regarding the variability of the negro as compared with the white. Ferguson[26] says that no certain conclusion as to racial variability can be reached, adding "It may finally be remarked with regard to the relative variability of whites and negroes, that it would not be at all surprising if groups of so-called negroes were definitely shown to be more variable than comparable groups of whites. For the groups that are generally called negroes are composed of in-

[23] Downey, J. E. *The Adolescent Will-Profile.* Journal of Educational Psychology, Vol. 11, No. 3, March, 1920.

[24] Bevis, W. M. *Psychological Traits of the Southern Negro.* American Journal of Psychiatry, Vol. 11, No. 1, July, 1921.

[25] Odum, H. W. *Social and Mental Traits of the Negro.* Columbia University Studies in History, Economics and Public Law, Vol. 37, No. 3, Whole No. 99, 1910.

[26] Ferguson, G. O., Jr. *The Psychology of the Negro.* Archives of Psychology, Monograph No. 36, April, 1916.

FIGURE 3.
Distribution of Individual Scores in the Different Traits.
————Negroes - - - - Whites

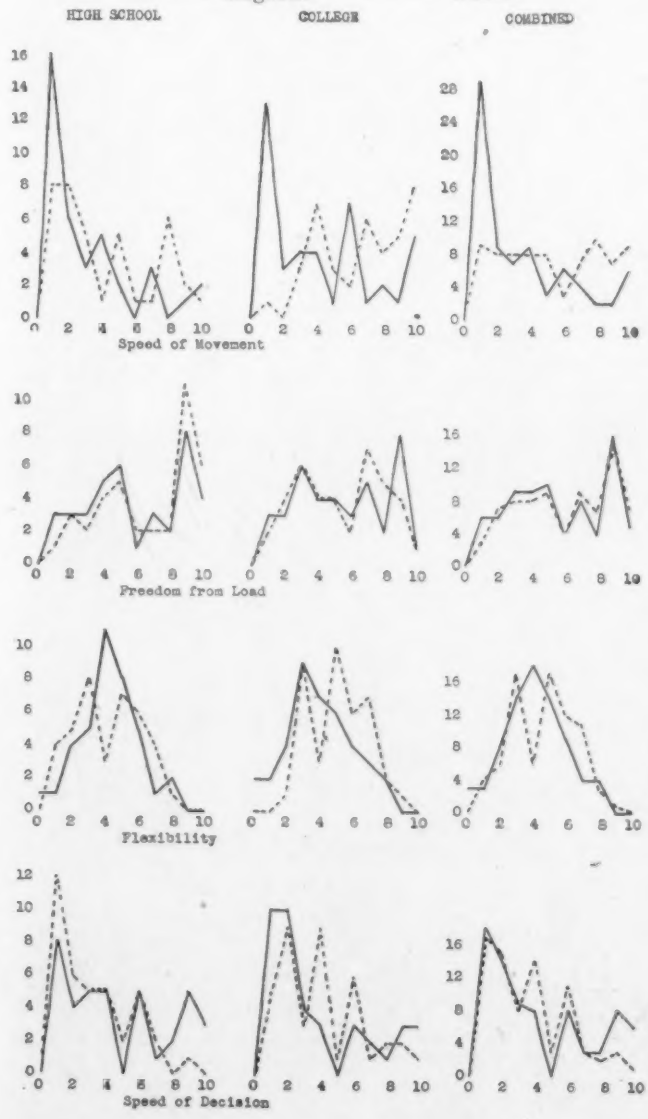


FIGURE 3. (cont.)

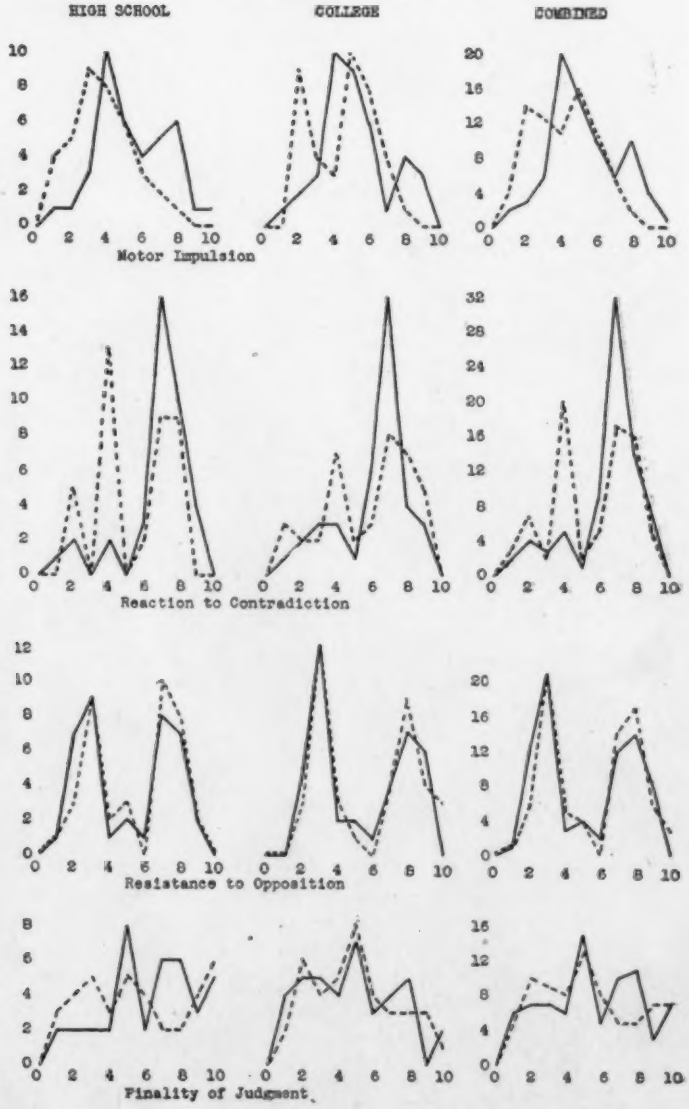
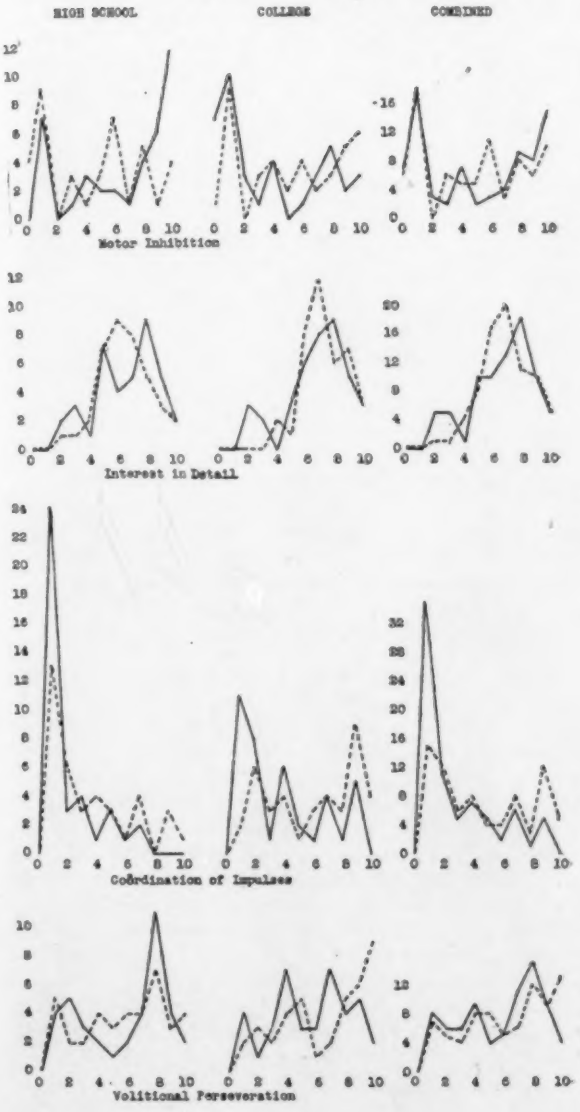


FIGURE 3. (cont.)



dividuals ranging from pure negroes to persons almost white, and it would be reasonable to suppose that such groups would vary more than would homogeneous white groups." Mayo[27] states that mental variability is less in negro high school students than in white high school students, the ratio being 93 to 100. Derrick[28]

TABLE II.

Comparison of the Standard Deviations of the Scores Made by the White and Negro College Students in the Various Traits

| Trait | White S. D. | Negro S. D. | White Excess | White Per Cent | Negro Excess | Negro Per Cent |
|---|---|---|---|---|---|---|
| Speed of Movement | 2.531 | 3.070 | | | .539 | 21.3 |
| Freedom from Load | 2.537 | 2.716 | | | .179 | 7.0 |
| Flexibility | 1.740 | 2.000 | | | .260 | 14.9 |
| Speed of Decision | 2.888 | 3.045 | | | .157 | 5.4 |
| Motor Impulsion | 1.781 | 1.955 | | | .174 | 9.7 |
| Reaction to Contradiction | 2.451 | 1.997 | .454 | 22.2 | | |
| Resistance to Opposition | 2.731 | 2.641 | .090 | 3.4 | | |
| Finality of Judgment | 2.392 | 2.301 | .091 | 3.8 | | |
| Motor Inhibition | 3.409 | 3.519 | | | .110 | 3.2 |
| Interest in Detail | 1.489 | 2.146 | | | .657 | 44.1 |
| Co-ordination of Impulses | 2.882 | 2.838 | .044 | 1.5 | | |
| Volitional Perseveration | 2.965 | 2.524 | .441 | 18.2 | | |
| Average Per Centage of Excess | | | | 9.8 | | 15.1 |

finds the opposite in regard to college students, when compared in respect to intelligence. He says "The white college student was about 77 per cent as variable as the negro." Nor is anything definite concerning variability found in the results from the Will-Temperament Tests. Inspection of the distribution curves for the scores in the various tests (Figure 3) shows practically no preponderance of either race. Comparison of the standard deviations from the mean in the scores made by college students, Table II, shows that the negro deviations exceed those of the whites in seven of the traits, and that the amount of excess is relatively greater. There would, then, seem to be a slight but not striking tendency for the negro to be more variable.

After the comparisons given above were made, the results were analyzed in terms of ages and academic grades. Except for the preponderance of freshmen in the white high school group, and

[27] Mayo, M. J. *The Mental Capacity of the American Negro.* Archives of Psychology, Monograph No. 28, 1913.

[28] Derrick, S. M. *A Study of the Intelligence of White and Colored College Students.* Journal of Applied Psychology, Vol. 4, pp. 316-329, December, 1920.

## TABLE III.

Distribution of subjects by ages and grades.

WHITE STUDENTS

| | High School | | | | College | | | |
|---|---|---|---|---|---|---|---|---|
| Age | Fresh. | Soph. | Jun. | Sen. | Fresh. | Soph. | Jun. | Sen. |
| 10 | 1 | | | | | | | |
| 11 | | | | | | | | |
| 12 | | | | | | | | |
| 13 | 5 | | | | | | | |
| 14 | 2 | 1 | 1 | | | | | |
| 15 | 5 | 1 | 1 | | | | | |
| 16 | 3 | | | 1 | | | | |
| 17 | 1 | 1 | | 2 | 1 | | | |
| 18 | 2 | 1 | 1 | 1 | 2 | 5 | | |
| 19 | 1 | | | | 1 | 6 | 3 | 1 |
| 20 | 1 | 1 | | | | 5 | 3 | |
| 21 | | 2 | 1 | | | 3 | 4 | |
| 22 | | 1 | | | | 2 | | |
| 23 | 1 | | | | | 1 | | |
| 24 | | | | | | | 1 | |
| 25 | | | | | | | | |
| 26 | | | | | | 1 | | |
| Total | 22 | 8 | 4 | 4 | 4 | 23 | 11 | 1 |

NEGRO STUDENTS

| | High School | | | | College | | | |
|---|---|---|---|---|---|---|---|---|
| Age | Fresh. | Soph. | Jun. | Sen. | Fresh. | Soph. | Jun. | Sen. |
| 12 | 1 | | | | | | | |
| 13 | | | | | | | | |
| 14 | 2 | | | | | | | |
| 15 | | 1 | | | | | | |
| 16 | 1 | | | | | | | |
| 17 | 2 | | | | 1 | | | |
| 18 | 3 | 5 | 1 | 1 | 1 | | | |
| 19 | 2 | 2 | 1 | 1 | 1 | 2 | | |
| 20 | | 1 | 2 | 1 | 3 | 2 | | |
| 21 | | | 1 | 1 | 2 | | 2 | 1 |
| 22 | | | | | | 2 | 1 | 1 |
| 23 | 1 | | 1 | 3 | 1 | 1 | | 2 |
| 24 | | | 2 | | 1 | 1 | 2 | 2 |
| 25 | | | 1 | | | 1 | 1 | |
| 26 | | | | 1 | | | 1 | 2 |
| 27 | | | | | 1 | | | 3 |
| 28 | | | | | | | | 1 |
| Total | 12 | 9 | 9 | 8 | 11 | 9 | 7 | 12 |

## TABLE IV.

Mean scores by ages of subjects.

WHITE STUDENTS

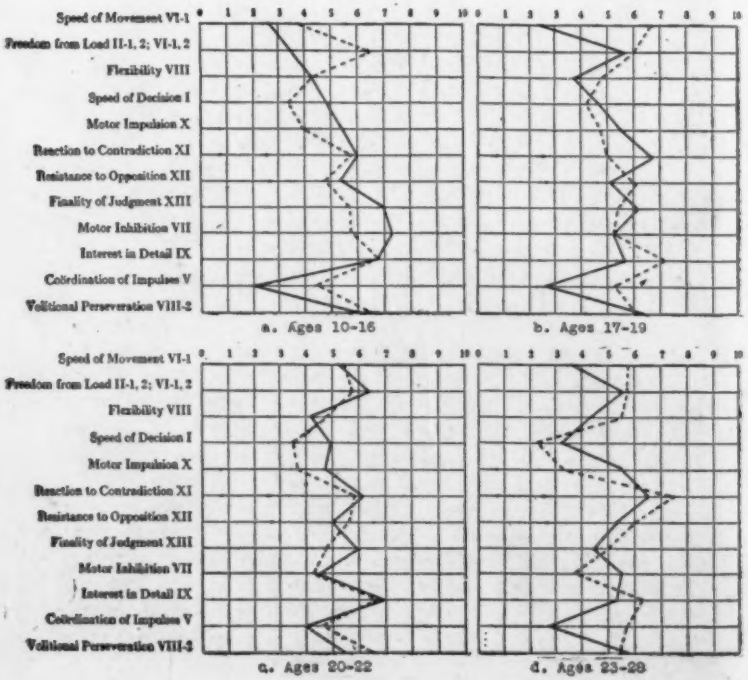
| Age | SM | FL | F | SD | Imp. | RC | RO | FJ | Inh. | ID | CI | VP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10-16 | 3.81 | 6.38 | 4.24 | 3.38 | 3.90 | 5.81 | 4.81 | 5.66 | 5.76 | 6.76 | 4.38 | 6.28 |
| 17-19 | 6.65 | 6.07 | 4.89 | 4.20 | 4.79 | 5.00 | 6.14 | 5.38 | 5.24 | 7.14 | 5.20 | 6.07 |
| 20-22 | 5.43 | 5.82 | 4.78 | 3.48 | 3.69 | 5.91 | 5.69 | 4.78 | 4.17 | 6.82 | 4.56 | 6.52 |
| 23-28 | 5.75 | 5.75 | 5.50 | 2.25 | 3.25 | 7.50 | 6.00 | 5.00 | 3.75 | 6.25 | 5.75 | 5.50 |

NEGRO STUDENTS

| Age | SM | FL | F | SD | Imp. | RC | RO | FJ | Inh. | ID | CI | VP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10-16 | 2.60 | 3.40 | 4.20 | 4.80 | 5.40 | 6.00 | 5.40 | 7.00 | 7.20 | 6.80 | 2.00 | 6.00 |
| 17-19 | 2.35 | 5.74 | 3.87 | 4.65 | 5.49 | 6.69 | 5.17 | 6.09 | 5.26 | 5.65 | 2.69 | 6.22 |
| 20-22 | 5.15 | 6.30 | 4.10 | 5.00 | 4.85 | 6.10 | 5.00 | 5.95 | 4.50 | 6.95 | 4.00 | 5.40 |
| 23-28 | 3.62 | 5.58 | 4.31 | 3.24 | 5.55 | 6.48 | 5.31 | 4.55 | 5.52 | 5.34 | 2.76 | 5.79 |

For number of subjects in each group, and grades represented, see Table 3.

## Figure IV.



of sophomores in the white college group, the distribution of students is approximately the same in both races. Table III shows the distribution by age and grades. From inspection of this table, it is seen at once that the negro student is from two to three years older than the white student of the same grade. This retardation may be caused by adverse environment or by inferior intelligence of the negro. Probably both factors are at work. To eliminate as much as possible the influence of age on the scores made by the groups, the subjects were divided into four sections, the first section containing those 16 years of age and younger, the second section containing those from 17 through 19, the third those from 20 through 22, and the fourth those 23 and above. Table IV gives the mean scores made by each section in each trait and Figure 4 shows the profiles plotted from these data. It will be seen from inspection that, as the subjects grow older, there is a tendency for increased scores in speed of movement, flexibility, reaction to contradiction, and co-ordination of impulses; a tendency for decreased scores in speed of decision, finality of judgment, motor inhibition, and perhaps volitional perseveration; the scores for motor impulsion tend to remain about the same; in freedom from load, the white scores remain about the same, or decrease slightly, while the negro scores increase with age; in resistance to opposition, the white scores increase, and the negro scores remain about the same; in interest in detail, the white scores decrease slightly, while the negro scores vary. The increase in speed of movement may be due to increased use of writing, the necessity for rapid note-taking, or the influence of other situations in every-day school life that call for speed in writing. The increase in the three other traits noted would probably be due to a deeper realization of problems as such, and the consequent tendency to eliminate "snap judgments". The decrease in motor inhibition and volitional perseveration may be due to the greater value which the older person puts upon time, or, in the case of motor inhibition, the decrease of content which the test-phrase holds for him, and the consequent increase in difficulty of the test. A person's rating in motor impulsion is probably a fundamental part of his make-up, and would hardly be expected to change in either direction. A true race difference may be indicated by the difference in the change of scores for freedom from load, in which those for the whites remained nearly constant, and those for the negroes increased. The negro score in this trait never increased enough, however, to bring it up to the lowest score of the whites. In such a trait as resistance to opposition, the increase in white scores, while the negro scores remain nearly constant, may indicate an earlier maturity on the part of the

negro. The decrease in scores for the white students in interest in detail may indicate, as noted above, an increased valuation of time.

In order to check the above classification of the groups into age-sections, another classification was made, this time into grade sections. Here it is evident, the factors of intelligence and greater capability come into play, and we may be fairly sure that the higher the grade, the higher the intelligence level, since there is a selective factor at work to eliminate those of less native capacity. Table V gives the mean scores attained by the various

TABLE V.

Mean scores by school grades of subjects.

WHITE STUDENTS

| Grade | SM | FL | F | SD | Imp. | RC | RO | FJ | Inh. | ID | CI | VP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HS 1-2 | 3.56 | 6.46 | 3.93 | 2.76 | 3.70 | 5.60 | 5.50 | 5.36 | 4.36 | 6.16 | 3.90 | 5.66 |
| HS 3-4 | 6.25 | 7.62 | 4.87 | 4.87 | 4.00 | 5.12 | 5.50 | 6.50 | 5.50 | 7.50 | 3.62 | 5.85 |
| C 1-2 | 7.04 | 5.74 | 5.04 | 4.00 | 4.33 | 6.89 | 5.74 | 4.88 | 6.96 | 7.22 | 6.44 | 6.58 |
| C 3-4 | 6.08 | 7.66 | 5.41 | 4.33 | 4.91 | 4.83 | 5.83 | 5.00 | 4.17 | 7.58 | 5.08 | 6.75 |

NEGRO STUDENTS

| Grade | SM | FL | F | SD | Imp. | RC | RO | FJ | Inh. | ID | CI | VP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HS 1-2 | 2.09 | 5.09 | 3.38 | 4.33 | 5.57 | 6.47 | 5.00 | 6.81 | 6.38 | 4.91 | 2.33 | 5.71 |
| HS 3-4 | 4.06 | 6.82 | 4.23 | 5.18 | 5.35 | 7.23 | 5.12 | 5.65 | 7.23 | 7.29 | 1.88 | 5.88 |
| C 1-2 | 4.40 | 5.25 | 4.05 | 4.75 | 4.95 | 5.85 | 5.65 | 5.20 | 5.45 | 6.15 | 3.85 | 5.30 |
| C 3-4 | 3.85 | 5.79 | 3.95 | 2.95 | 4.84 | 6.37 | 5.42 | 4.21 | 3.79 | 7.58 | 3.89 | 6.10 |

Average ages of above groups:

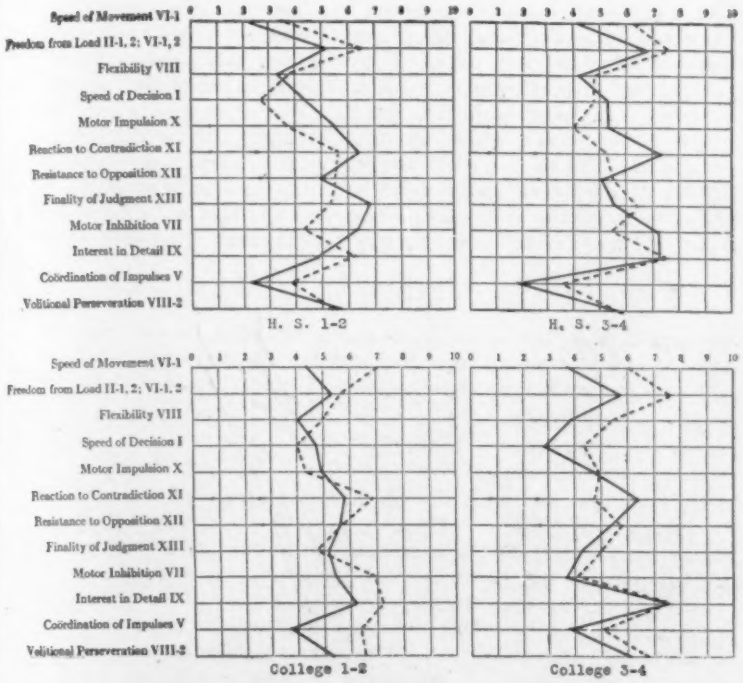
| | HS 1-2 | HS 3-4 | C 1-2 | C 3-4 |
|---|---|---|---|---|
| Whites | 16.3 | 17.0 | 19.74 | 20.33 |
| Negroes | 17.66 | 21.59 | 21.2 | 25.58 |

For distribution of ages, and numbers in each group, see Table 3.

sections, and Figure 5 shows the profiles plotted from the data. In this case, with increase in grade, there is an increase in the scores of more traits than when the student's increase is in age. Inspection shows that there are increases in the scores for speed of movement, flexibility, resistance to opposition, interest in detail, co-ordination of impulses and volitional perseveration. Finality of judgment is the only trait in which there is a decrease. The scores for freedom from load, reaction to contradiction, and motor inhibition remain about the same, or vary irregularly. There is a slight increase in the scores for speed of decision and motor impulsion for the whites, with a decrease in these traits for the negroes. With increased intelligence and stronger personality, one would naturally expect higher scores. For finality of judgment,

the decrease is from above "5" to about "5" which is the normal score. The advanced students seem to be less addicted to hasty revision. The constancy of the traits freedom from load, reaction to contradiction, and motor inhibition, may indicate that normal scores in these traits are sufficient for proficiency in school work,

Figure V




although it would seem that the advanced student would gain in resistance to contradiction. The opposite direction in which the scores tend, for the two races, in speed of decision and motor impulsion is probably indicative of racial characteristics, and may offer good returns for further investigation.

Since, as stated in the first part of this section, a study of mean or median scores offers little that is definite in the way of racial temperament patterns, the scores for the college students were gone over in an attempt to list any tendency toward groupings of the high scores found there. Following Professor Downey in the

main, we may suggest that there are three types of such groupings—those in the traits listed as the first four characterizing one as mobile or rapid-fire in organization, as the second four characterizing him as aggressive or forceful, and as the third four the controlled, deliberate, careful type. Also, there may be combinations of any two groupings, or of none at all—the profile may be balanced or else very irregular. The method followed here in listing tendencies was as follows: The first four scores in each individual's case were added together, as were the second and third sets of fours. These sums were compared, and when the sum of one group exceeded the sum of either of the other groups, the individual was listed as being characterized by the tendency correlated with that group. Where no one group exceeded either of the others by three or more points, there were two possibilities—either that the individual had a well balanced profile, or else there was a combination of two high groups and one low one. If the first was the case, he was listed as "Balanced"; if the second was the case, he was listed under the combination

TABLE VI

Showing distribution of college students according to the group of traits in which each received his highest scores.

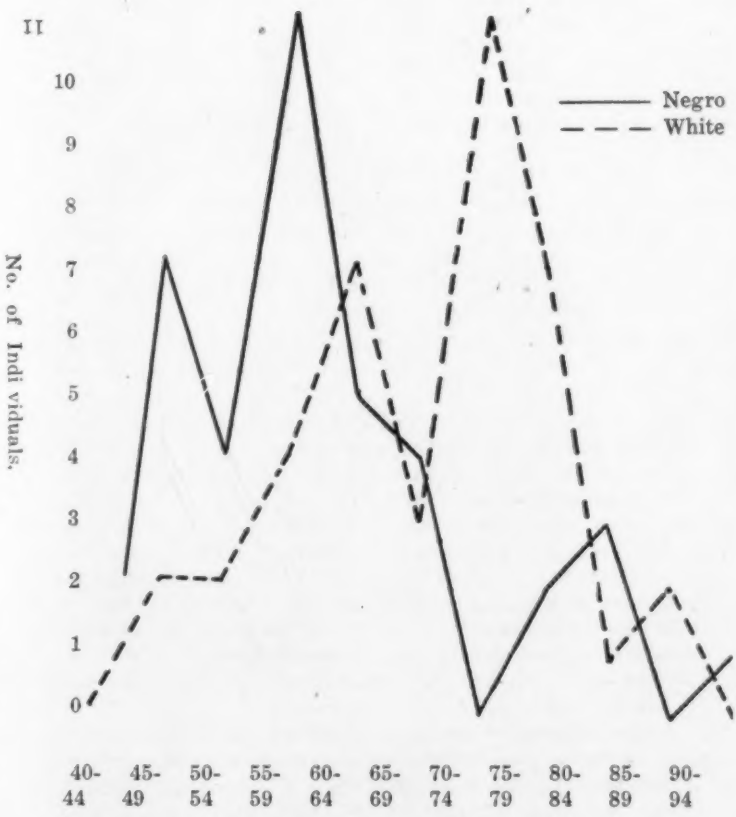
| | 1st | 2nd | 3rd | 1st and 2nd | 2nd and 3rd | 1st and 3rd | Balanced |
|---|---|---|---|---|---|---|---|
| Whites | 4 | 6 | 14 | 1 | 6 | 5 | 3 |
| Negroes | 2 | 9 | 8 | 4 | 8 | 3 | 5 |

of his two high groups. Table VI gives the results of this attempt to classify the patterns. From this table we may see that the whites slightly surpass the negroes in the number of mobile, rapid-fire individuals; they have a clear superiority in the number of controlled, deliberate, careful persons; and they slightly surpass the negroes in the number of individuals showing a combination of these two characteristics. The negroes slightly surpass the whites in the number of aggressive persons and in the number of persons combining quickness and mobility with aggressiveness, and also in the number of individuals combining aggressiveness and deliberation.

As a last point, Professor Downey suggests that, while there is not as much significance attached to the total score of each individual as to his scores on various traits, still, a high total score is, in general, indicative of a "stronger personality," and a low total score is indicative of a weaker personality. On this assumption, a tabulation of total scores of the college group was made, and a distribution curve was drawn from the results. See Figure 6.

FIGURE 6.

Distribution of total scores of College students.




| | Mean | Median |
|---|---|---|
| White | 68.08 | 70 |
| Negro | 59.05 | 57 |

Six negrces (15.4%) passed the white median.

It was found that the white college students surpassed the negro college students in their totals, which, therefore, would indicate, that, in general, the whites have "greater force of personality."

## SUMMARY

1. In general, the negro is slower in movement than the white, has about the same or slightly greater load or inertia; has slightly less flexibility; is slightly quicker in making his decisions; his motor impulsion is slightly greater; reacts more firmly against contradiction; but offers less resistance to physical opposition; takes about the same or slightly less time in reconsidering his decisions; the negro college student has less motor inhibition, the negro high school student has greater, compared with white students; the negro has practically the same interest in detail as the white, has markedly less ability in co-ordinating his impulses; and has a slightly less tendency to volitional perseveration.

2. There is little difference in variability, but the negro college student has greater variability on seven of the traits, as opposed to the greater variability of the whites in five, and the amount of the negroes' variability is slightly greater.

3. In general, the negro college and high school student is from two to three years older than the white student of the same academic grade.

4. As to the influence of age (from early adolescence to the middle twenties, as the subject approaches maturity, there is a tendency for increased scores in speed of movement, flexibility, reaction to contradiction, and co-ordination of impulses; a tendency for decreased scores in speed of decision, finality of judgment motor inhibition, and volitional perseveration, the scores for motor impulsion tend to remain about the same; in freedom from load, the white scores remain about the same, or decrease slightly, and the negro scores increase with age; in resistance to opposition, the white scores increase and the negro scores remain about the same; in interest in detail, the white scores decrease, while the negro scores vary.

5. As to the analysis of mean scores by grades (probably influenced by the factor of increasing intelligence in direct ratio to the higher grades): the scores for speed of movement, flexibility, resistance to opposition, interest in detail, co-ordination of impulses, and volitional perseveration increase in the higher grades; the scores for freedom from load, reaction to contradiction, and motor inhibition vary; there is a slight increase in the scores for speed of decision and motor impulsion for the whites, with a decrease in these traits for the negroes, from the lower to the higher grades.

6. In predominance of temperamental patterns, the whites slightly surpass the negroes in the number of mobile, rapid-fire individuals; they have a clear superiority in the number of con-

trolled, deliberate, careful persons; and they slightly surpass the negroes in the number of individuals showing a combination of these two characteristics. The negroes slightly surpass the whites in the number of aggressive persons and in the number of individuals combining quickness and mobility with aggressiveness, and also in the number of individuals combining aggressiveness and deliberation.

7. In "strength of personality," as indicated by the total score of each individual, the whites surpass the negroes, only 15.4 per cent of the negroes exceeding the median of the whites.

# A STUDY OF THE DATA ON THE RESULTS GATHERED FROM REPEATED MENTAL EXAMINATIONS OF 200 DEFECTIVE CHILDREN ATTENDING SPECIAL SCHOOLS OVER A PERIOD OF EIGHT YEARS

By META ANDERSON

Director Binet Classes, Newark, N. J.

Although the work for defectives was begun in Newark in September, 1910, it was not until 1913 that the organization of the work permitted anything like systematic re-examination of the children admitted to the special schools and even then complete yearly re-examination of the whole special school population was not accomplished. However, since 1916 it has been possible to re-examine each child in the special schools yearly excepting a few cases which will be noted later. The object of the work was to get information in regard to the growth of intelligence of the children in the special classes (now called Binet Classes) in order, if possible, to know better how to train them and to eliminate errors in their training and education. The object of this study was to organize, classify and analyze these data in order to get further information in regard to types of mental growth curves and in order to get further information in regard to the value of the intelligence quotient.

While the population of the Binet Schools is upward of five hundred pupils only two hundred were studied here. These two hundred cases vary from the grade of low grade imbecile to that of high grade moron. Before January, 1918, the cases were selected for the Binet schools because they were failures in the regular grades and at the same time fell more than three years below normal according to an intelligence test (this standard being set by the New Jersey law). Since January, 1918, the cases have been selected by the Psycho-Educational Clinic of the city. Each case is certified as being definitely feeble-minded before being sent to a Binet School.

Each case was examined from three to eight times at not less than a year's interval. The majority of the cases were re-examined by the writer. The others were examined by two head assistants in the department both of whom were unusually well trained to do the work.

In a department always having a population of 500 defective children it would seem possible to have numbers of retests on at least that many children. This investigation has shown that the turnover in the population of the Binet Schools is very great.

Many of the children except the lowest grade are assigned to special work late in their school career. These children are not presented for psychological examination until the regular schools have done all in their power to help them succeed in their work. Hence, it is only after their inability to learn has become most marked or their behavior disorders have become unbearable that these children are assigned to the Binet Schools. By that time the children are anywhere from ten to fourteen years of age. Since they leave school at sixteen and sometimes earlier as during the war when the pressure for workers was very great these children have short opportunity for training. This also renders it impossible to have a great number of retests on the same individual.

A certain percentage of this group also leave school to go to houses of correction, to institutions, move from the city or are lost track of. Therefore the number remaining in the Binet Schools for more than four years is comparatively few.

With the present wider use of group tests defective children are being discovered earlier in their school career but it will be necessarily a number of years before sufficient data are gathered on this new group for a worthwhile study.

Before 1918 the Goddard revision of the Binet tests were used, since 1918 the Stanford revision of the Binet tests have been used.

The study falls into two parts. The first part has to do with the frequency of types of mental growth curves and the second part has to do with a study in values of the IQ.

## PART I. FREQUENCY OF TYPES OF MENTAL GROWTH CURVES

### Other Investigations.

Kuhlman has made growth curves on data gathered from repeated retests on inmates of the Minnesota institution and he says: "The main features of these growth curves present but little that is not now pretty well known. They show that the feeble-minded do develop mentally, and at a rate in proportion to the grade of intelligence. This was still a disputed point at the time this study was begun. This rate of development decreases with age for all grades. The rate of this decrease cannot, of course, be shown with this data, as we do not know how much mental growth in terms of absolute units the mental year represents at different age levels."

Kuhlman summarizes his results on mental age growth as follows:

"1. The mental ages of the feeble-minded increase with age at a rate proportionate to the degree of mental deficiency.

"2. On the whole the mental age ceases to increase between the ages of fifteen to eighteen, the idiot grade ceasing development about three years earlier than the borderline grade.

"3. For the cases figuring in the present study and for the time period they were under observation, 4.8 per cent of the cases gained twelve months or more in mental age in a year, 68 per cent gain from one to eleven months a year, 11 per cent gain or lose less than one month a year, and 16 per cent lose from one to seven months a year in mental age.

"4. The lower grades lose more frequently in mental age than the higher grades.

"5. The frequency in loss in mental age increases with age, independently of grade."[1]

Doll says on the rate of growth for subnormals: "Not much more is known about the rate of mental growth among extreme subnormals, like the feeble-minded, than about the age of arrest. In this regard we are even at a loss for adequate observational data. Experience with institutional feeble-minded, however, suggests a great variety of possible forms of retardation curves. Such experience points to a small but significant group of feeble-minded who grade 'at age' by the Binet scale when first examined but who make no material progress in later examinations over a period of years (Type A.) There are others (Type B) who at the time of first examination test at age, then continue to develop at a retarded rate for a time and then reach a point of arrest. There is a third group (Type C) already retarded when first observed, who from the first examination to the age of final arrest develop at an annual rate of growth which is a constantly decreasing variable. There is also a fourth group (Type D) who from the time of first examination to the age of final arrest. There are many variants and combinations of these basic types for individual subjects. We believe, however, that these four types comprise the majority of the feeble-minded. It is conceivable, though not probable, that some individuals might have a higher than average rate before retarding. It is also conceivable, though not likely, that the arrest might be relatively greater in the early ages with a curve of growth which would be accelerating at some intervals. It remains for experimental investigation to show the relative frequencies of these different types of growth curves."[2]

[1]Kuhlmann, Results of Repeated Re-examinations of Feeble-minded (639) Over a Period og Ten Years. Journal of Applied Psy. Sept., 1921.
[2]Doll, Growth of Intelligence. Psy. Monograph, pp. 15-19.

With the view of determining what light, if any, the limited data on hand could throw on the subject of the growth of intelligence of feeble-minded subjects 100 mental age growth curves of pupils attending the Binet Schools from three to eight years were plotted. These curves were then classified according to Doll's theoretical intelligence growth curves as follows:

*Type A.* (Doll) Those who test "at age" by the Binet scale when first examined but make no material progress in later examinations.

*Type B.* (Doll) Those who at the time of first examination test "at age," then continue to develop at a retarded rate for a time and then reach a point of arrest.

*Type C.* (Doll) Those who are already retarded at time of first examination but who continue to develop from the time of first examination until time of final arrest at an annual rate of growth which is a constantly decreasing variable.

*Type D.* (Doll) Those who from the time of first examination to the age of final arrest develop at a constant rate up to time of final arrest.

For convenience two new types were added although it is quite likely they could be easily included in one of the above groups. They are as follows:

*Type E.* (New) Those who are retarded at the time of first examination, and who also have apparently reached the time of final arrest.

*Type F.* (New) Those who have developed at a constant or variable rate up to the time of apparent arrest, but who after two to four years of apparent arrest, have begun to develop again.

| NUMBER OF CASES | TYPE |
|---|---|
| 1 | 4 to 5 |
| 1 | A |
| 0 | B |
| 47 | C |
| 22 | D |
| 19 | E |
| 11 | F |
| 100 | |

Of the 11 cases listed as *type f* there were

4 who remained at an apparent arrest for 2 years.
5 who remained at an apparent arrest for 3 years.
2 who remained at an apparent arrest for 4 years.

Of these 11 cases listed as *type f* there were

2 cases who continued development for 1 year after arrest.
6 cases who continued development for 2 years after arrest.
2 cases who continued development for 3 years after arrest.
1 case who continued development for 4 years after arrest.

at which time there was another apparent arrest or the children left school.

*Type f cases classified according to life age*

| Number of Cases | Life Age |
|---|---|
| 1 | 6 to 7 |
| 1 | 7 to 8 |
| 5 | 9 to 10 |
| 1 | 10 to 11 |
| 2 | 11 to 12 |
| 1 | 12 to 13 |
| 11 | |

*Type f cases classified according to mental age*

| Number of Cases | Mental Age |
|---|---|
| 2 | 3 to 4 |
| 6 | 6 to 7 |
| 1 | 7 to 8 |
| 1 | 8 to 9 |
| 11 | |

## SUMMARY.

1. It is evident that a greater number (47) of the cases studied in this group have developed mentally according to Doll's theoretical curve described as Type C.

2. The next larger group (22) of cases studied have developed mentally according to Doll's theoretical curve described as Type D.

3. There were 67 cases out of the 100 studied who were retarded when first examined and who then retarded either at a constant rate or at a steadily decreasing rate up to age of final arrest.

4. Only one (1) case could be obviously classified under Type A.

5. The 19 cases which were retarded when first examined and which developed no further at later examinations may or may not belong to Type A.That is, it is possible that if they had been examined at earlier periods they would have tested "at age."

6. The 11 cases under Type F could be classified under Type C in a broad classification. They differ from that type merely in the fact of having long periods of arrest but in the long run they develop at a decreasing rate up to the time of final arrest.

7. It would appear from the study of these 100 cases that the majority 80 in this study of feeble-minded cases slow up gradually in mentality up to time of final arrest and that comparatively few develop normally for a time and then stop suddenly. In this study not more than 20 of the cases could have been said to have come to sudden arrest and 19 out of the 20 cases are doubtful.

8. The 100 cases studied were all classified as feeble-minded and were attending Binet Schools. It is possible that feeble-minded cases which develop normally and then retard suddenly are not discovered early enough in their school careers to be selected for Binet Schools.

## Part II. Study in Some Values of the I. Q.

From three to seven tests were made on 200 children attending the Binet Schools at intervals of a year in most cases. There are cases which are noted in which the retests were made at greater intervals because the children were not available for examination either because they had moved from the city or attended another school and then re-entered the Binet School when they were re-examined. At the time of this study there were only two hundred cases who had been re-examined a sufficient number of times to make the study of their cases worth while. These cases then were all selected because they were attending the Binet Schools and because they had attended long enough to have from two to six retests made at intervals of a year.

The cases were classified and grouped.

(a) According to life age at initial test and I. Q. in groups of tens.

(b) According to I. Q. in groups of tens.

(c) According to I. Q. and intervals of retests regardless of life age.

These groupings were made in order to study the I. Q. in its various aspects.

A. CASES CLASSIFIED ACCORDING TO LIFE AGE AT INITIAL TEST AND I Q—IN GROUPS OF TENS.

| No. of Cases. | Life Age. | Median I. Q. at First Test. | Number of Retests. | Median I. Q. at Last Test. |
|---|---|---|---|---|
| 0 | 6 | | | |
| 2 | 7 | 38½ | 7 | 44½ |
| 1 | 8 | 39 | 6 | 41 |
| 1 | 9 | 33 | 4 | 44 |
| 2 | 10 | 31 | 4 | 25 |
| 0 | 11 | | | |
| 0 | 12 | | | |
| 1 | 13 | 36 | 4 | 29 |
| 7 | | | | |
| 1 | 6 | 44 | 6 | 42 |
| 0 | 7 | | | |
| 3 | 8 | 42 | 6 | 44 |
| 3 | 9 | 47 | 4 | 47 |
| 0 | 10 | | | |
| 2 | 11 | 44 | 3 | 48 |
| 1 | 12 | 49 | 5 | 40 |
| 0 | 13 | | | |
| 10 | | | | |
| 3 | 6 | 55 | 4 | 44 |
| 1 | 7 | 58 | 5 | 58 |
| 3 | 8 | 59 | 5 | 55 |
| 7 | 9 | 54 | 3 | 49½ |
| 4 | 10 | 53½ | 3 | 51½ |
| 6 | 11 | 54½ | 4 | 51 |
| 2 | 12 | 57 | 3 | 48½ |
| 26 | | | | |
| 0 | 6 | | | |
| 1 | 7 | 69 | 3 | 57 |
| 3 | 8 | 61 | 4 | 53 |
| 15 | 9 | 67 | 5 | 57 |
| 5 | 10 | 67 | 4 | 58 |
| 8 | 11 | 64½ | 3 | 57 |
| 6 | 12 | 64 | 3 | 57½ |
| 2 | 13 | 64 | 3 | 56 |
| 40 | | | | |

A. CASES CLASSIFIED ACCORDING TO LIFE AGE AT INITIAL TEST AND I Q—IN GROUPS OF TENS. (CONTINUED)

| No. of Cases. | Life Age. | Median I. Q. at First Test. | Number of Retests. | Median I. Q. at Last Test. |
|---|---|---|---|---|
| 2 | 6 | 74 | 4 | 60½ |
| 1 | 7 | 70 | 7 | 60 |
| 2 | 8 | 75 | 4 | 65 |
| 6 | 9 | 74 | 6 | 48 |
| 5 | 10 | 73½ | 5 | 56½ |
| 8 | 11 | 73½ | 3 | 60 |
| 5 | 12 | 72 | 3 | 60 |
| 29 | | | | |

## SUMMARY OF RESULTS.

1. The cases with I. Q.'s from 30 to 39, inclusive, and with life ages from six to nine years, inclusive, showed a gain in median I. Q. of 2 to 11 points.

2. The cases with I. Q.'s from 30 to 39, inclusive, and with life ages from 10 to 13, inclusive, showed a loss of 6 to 7 points in median I. Q.

3. The cases with I. Q.'s from 40 to 49, inclusive, and with life ages of 6 years showed a loss of 2 points in median I. Q.

4. The cases with I. Q.'s from 40 to 49, inclusive, and with life ages from 8 to 11, inclusive, showed a gain in median I. Q. of 2 to 4 points.

5. The cases with I. Q.'s from 40 to 49, inclusive, and with the life ages of 12 showed a loss in median I. Q. of 5 points.

6. The cases with I. Q.'s from 50 to 59, inclusive, and with life ages from 6 to 12, inclusive, showed a loss in median I. Q. of nothing to 9 points.

7. The cases with I. Q.'s from 60 to 69, inclusive, and with life ages from 6 to 13, inclusive, showed a loss in median I. Q. of 7 to 12 points. The cases with I. Q.'s between 40 and 49 showed a loss of 1 to 2 points.

8. The cases with I. Q.'s from 70 to 79, inclusive, and with life ages from 6 to 12, inclusive, showed a loss in median I. Q. of 10 to 26 points.

9. In other words, the younger cases with I. Q.'s from 30 to 39, inclusive, gained in median I. Q.; the older cases with I. Q.'s from 30 to 39, inclusive, lost in median I. Q.

10. With the exception of the 6-year-olds who lost 2 points in median I. Q. the younger cases with I. Q.'s from 40 to 49, inclusive,

gained in median I. Q.; the older cases with I. Q.'s from 40 to 49, inclusive lost in median I. Q.'s.

11. All cases of any age with I. Q.'s from 50 to 80, inclusive showed a loss in median I. Q.

B. CASES CLASSIFIED ACCORDING TO IQ

| No. of Cases | I. Q. Range | Median I.Q.'s at Successive Examinations. 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 14 | 30-39 | 35 | 33 | 31 | 30 | 37 | |
| 29 | 40-49 | 47 | 49 | 46[2] | 48 | 46 | |
| 63 | 50-59 | 56 | 56 | 54 | 51[2] | 50 | 48 |
| 59 | 60-69 | 65 | 60 | 60 | 58 | 60 | 57 |
| 33 | 70-79 | 73 | 67 | 62 | 60 | 58 | 56 |

## SUMMARY.

1. The cases with I. Q.'s between 30 and 39 showed a gain of 2 points in median I. Q. at the 5th re-examination.
point in median I. Q. at the 5th re-examination.

3. The cases with I. Q.'s between 50 and 59 showed a loss in median I. Q. of 8 points at the 6th re-examination.

4. The cases with I. Q.'s between 60 and 69 showed a loss in median I. Q. of 8 points at the 6th re-examination.

5. The cases with I. Q.'s between 70 and 79 showed a loss in median I. Q. of 17 points at the 6th re-examination.

6. All cases with I. Q.'s below 50 remained practically stationary with a gain or loss of but 1 or 2 points in median I. Q.

7. The cases with I. Q.'s between 70 and 80 showed loss. The highest initial I. Q. showed the greatest loss in median I. Q. at the final re-examination.

C. CASES CLASSIFIED ACCORDING TO IQ AND ALSO ACCORDING TO INTERVALS OF RETESTS.

| No. of cases | IQ range | Median IQ's at initial test and at intervals of retests Compared. | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 12 | 30-39 | 34-32[2] | | | | | |
| 10 | 30-39 | | 33[2]-28[2] | | | | |
| 12 | 30-39 | | | 34-33 | | | |
| 9 | 30-39 | | | | 34-37 | | |
| 5 | 30-39 | | | | | 39-42 | |
| 33 | 40-49 | 46-48 | | | | | |
| 24 | 40-49 | | 46-45[3] | | | | |
| 20 | 40-49 | | | 44-48 | | | |
| 13 | 40-49 | | | | 44-42 | | |
| 7 | | | | | | 44-43 | |
| 67 | 50-59 | 56-56 | | | | | |
| 54 | 50-59 | | 55-53 | | | | |
| 40 | 50-59 | | | 55[2]-56[2] | | | |
| 19 | 50-59 | | | | 55-50 | | |
| 12 | 50-59 | | | | | 53[2]-48 | |
| 7 | 50-59 | | | | | | 56-49 |
| 74 | 60-69 | 65-61 | | | | | |
| 64 | 60-69 | | 65-60 | | | | |
| 43 | 60-69 | | | 65-58 | | | |
| 21 | 60-69 | | | | 67-60 | | |
| 9 | 60-69 | | | | | 65-56 | |
| 33 | 70-79 | 73-67 | | | | | |
| 31 | 70-79 | | 73-62 | | | | |
| 25 | 70-79 | | | 73-61 | | | |
| 16 | 70-79 | | | | 72-58 | | |
| 11 | 70-79 | | | | | 72-56 | |

## SUMMARY.

All median IQ's in all ranges showed a loss as follows:

| Range of IQ. | Points lost at successive examinations. | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 30-39 | 1 | 2 | 3 | 3 | 5 | |
| 40-49 | ½ | 1 | 2 | 2 | 4 | |
| 50-59 | 0 | 1 | 3 | 4½ | 5 | |
| 60-69 | 4 | 5 | 7 | 7 | 9 | |
| 70-79 | 6 | 8 | 11 | 14 | 16 | |

In each group of cases the greater loss in median IQ occurred at the end of the fifth examination.

In each group of cases the greatest loss in median IQ occurred in the groups of the higher initial IQ's.

## CONCLUSIONS.

1. A great majority of the cases who are examined as candidates for the Binet Schools are retarded at the time of first examination and then retard either constantly or at a decreasing ratio up to time of final arrest.

2. Comparatively few of the cases examined as candidates for the Binet Schools develop normally up to a certain age of arrest and then remain at arrest.

3. In cases whose Intelligence Quotient is below 50 there is a tendency for the median IQ to remain constant or to gain or lose slightly. The median IQ shows a gain among the younger cases.

4. All cases with IQ's between 50 and 80 show a decrease in median IQ at any age. The highest initial IQ's show the greatest loss in median IQ after repeated examinations.

5. In the case of examinations at varying intervals the greatest loss in median IQ was at the interval of five years which was the last examination for all grades of cases. The greatest loss in median IQ was shown by the highest grade of cases.

6. The IQ is variable in individual cases.

7. No group of cases showed a marked increase in median IQ.

8. These conclusions are merely for cases examined as candidates for Binet Classes. It is possible that repeated examinations on all retarded cases in a school system might give different results.

# THE DOWNEY WILL-TEMPERAMENT GROUP TEST; A FURTHER ANALYSIS OF ITS RELIABILITY AND VALIDITY

By G. M. RUCH and M. C. DEL MANZO,
University of Iowa.

This paper presents the results of a statistical analysis of the scores of one hundred and forty-six high school students tested with the *Downey Group Will-Temperament Test.* The discussion of these results will be divided into four sections in accordance with the specific objectives of the investigation.

## I THE RELIABILITY OF THE TEST AS A WHOLE

Since the Downey tests are published in only a single form the usual method of obtaining reliability coefficients by correlating equivalent forms is not applicable here. Moreover, the common practice of breaking tests into chance halves for purposes of obtaining the self-correlation can only be carried out with certain qualifications as to the justice of this method. In an earlier paper[1], one of the authors has argued that the correlation of the sums of the odd-numbered tests against the sums of the even-numbered tests might fairly provide an approximation to a valid coefficient of reliability. A few words in explanation of the nature of the *Downey Will-Temperament Tests* will help to make clear the limitations of the method of correlating chance halves.

The scale is composed of twelve separate tests each designed to measure one aspect of the total volitional pattern. These are:

A Speed of Movement (VI-1)
B Freedom from Load (VI-1, 2)
C Flexibility (VIII)
D Speed of Decision (I)

E Motor Impulsion (X)
F Self-Confidence (XI)
G Non-compliance (XII)
H Finality of Judgment (XIII)

I Motor Inhibition (VII)
J Interest in Detail (IX)
K Co-ordination of Impulses (V)
L Volitional Perseveration (VIII-2)

[1]Ruch, G. M.: "A Preliminary Study of the Correlations between Estimates of Volitional Traits and the Results from the Downey 'Will-Profiles'," Journal of Applied Psychology, 1921, 5, 159-162.

This arrangement of the twelve traits into three groups as here indicated follows the purpose of the author of the test in that traits A to D belong to a common "pattern"; traits E to H belong to a second "pattern"; and traits I to L belong to a third "pattern." The groupings are intended to throw together the tests that are intended to reveal, respectively, three main types of volitional temperaments known, in order, as:

I "Hair-trigger" organization, or the explosive type of will described by James (Tests A, B, C, and D).

II "Willful, aggressive pattern" (Tests E, F, G, and H).

III "Slow, accurate, tenacious type" (Tests I, J, K, and L).

The grouping and characterizations of the author of the tests is adhered to here strictly.

Since the scale is of the nature of a diagnostic picture in which certain groups (the three patterns) of tests can logically yield uniformly high scores while other portions can as logically yield very low scores (or at least much lower scores), it follows that the total score on the whole scale will have very little meaning. It also follows that any attempt to split the total scale into chance halves must take care that equal numbers of tests be taken from each of the three diagnostic patterns in forming such chance halves. If this is done properly we would then have two half-scales; each half-scale being composed of two tests from each of the three patterns or a total of six tests in each half scale. Such chance halves could be made up in this particular case in several way, viz., (1) the grouping of all of the odd-numbered tests in one half-scale and all of the even-numbered tests in the other half-scale, (2) the grouping of the first two tests of every pattern into one of the half-scales and the last two tests of each pattern into the other half-scale, or, (3) the grouping of the first and last tests of each pattern into one half-scale and the second and third tests of each pattern into the second half-scale.

In the paper mentioned above, the reliability of the Downey tests (individual form) was studied only by the first of these three methods. In the present report, the reliability coefficients have been determined by all of the three possible ways and the results averaged. Table 1 shows these results. The notations as "Method 1," "Method 2," and "Method 3" refer to the above-mentioned ways of obtaining chance halves. The notation "$r_{\frac{11}{22}}$" means that the coefficient of correlation stated was obtained by correlating half (six tests) of the scale with the remaining half (six tests) of the scale. The notation "$r_{12}$" is the porphecied correlation which would be obtained if two similar whole forms were correlated. The values of the "$r_{12}$" column were calculated by the

use of Brown's formula, $r = \frac{n\,r}{1+(n-1)\,r}$. The ages of the subjects cover a range of from ten to twenty-nine years. Grades seven to twelve are involved.

## TABLE 1

RELIABILITY COEFFICIENTS

| Method | $r_{\frac{1}{2}\frac{1}{2}}$ | $r_{11}$ (Brown's formula) | N |
|---|---|---|---|
| I | 0.17 | 0.29 | 146 |
| II | 0.31 | 0.47 | 146 |
| III | 0.24 | 0.38 | 146 |
| Average | 0.24 | 0.38 | 146 |

The probable errors of these coefficients (Pearson product-moment formula) range between 0.05 and 0.06. Table 7 shows the probable errors of coefficients of correlation for the various values of r reported in this study.

In view of the wide range of talent represented in the group tested, these reliability coefficients cannot be considered to be very satisfactory. They are, however, much higher than those published in the earlier paper cited; the range of talent in the former group being very much more curtailed.

There is another possibility which must be considered in accepting these reliability coefficients at their face value. According to our original assumptions, the reliability of the scale as a whole can only be obtained by the methods we have used *in case the tests are correctly grouped into patterns in a proper manner.* The low reliability which we have found might mean merely that the tests are improperly classified at present. In this case, the true reliability of the scale (properly arranged) might be markedly higher than we have reported. Discussion of this question of the grouping of the separate tests should be postponed until the intercorrelations of the tests have been presented.

## II The Internal Coherence of the Scale

One of the crucial issues in the evaluation of the Will-Temperament Test is the matter of the arrangement of the separate elements into the diagnostic patterns. As the authors interpret Dr. Downey's classification of the individual tests into groups which picture volitional patterns, it would seem that a valid grouping

would imply certain mutual relationships among the tests. These are:

(1) High scores in one test of a pattern would be paralleled with high scores in the other three tests of the same pattern. This is equivalent to saying that the inter-correlations between tests of a given pattern would tend to be positive and high.

(2) Correlations between each test and the composites of the tests of the same pattern would tend to be positive and high.

(3) The correlations between tests of one pattern and those of the other patterns would tend to be zero or very low.

(4) Correlations between single tests and the total score on the scale would be low in all probability and not significant for purposes of evaluating the tests.

In order to test these assumptions about the proper grouping of the tests of the present scale, the correlations of each test with the composite (sum) of the other three tests of the same pattern have been computed. For obvious reasons the test used as one of the variables in the correlation must be omitted from the composite. Table 2 shows the obtained correlations.

TABLE 2

CORRELATIONS OF THE TESTS WITH THE COMPOSITES

| Test | Composite | r |
|---|---|---|
| A | B+C+D | 0.54 |
| B | A+C+D | 0.28 |
| C | A+B+D | 0.10 |
| D | A+B+C | 0.39 |
| E | F+G+H | −0.08 |
| F | E+G+H | −0.23 |
| G | 1+F+H | −0.13 |
| H | E+F+G | −0.09 |
| I | J+K+L | 0.07 |
| J | I+K+L | 0.33 |
| K | I+J+L | 0.07 |
| L | I+J+K | 0.05 |

The average inter-correlations of the tests within a pattern is probably even a better method of testing the general validity of the grouping into patterns. Table 3 shows the results yielded by this method.

TABLE 3

INTER-CORRELATIONS OF THE TESTS OF THE SAME PATTERNS

| *Variable 1* | *Variable 2* | *r* |
|---|---|---|
| A | B | 0.36 |
| A | C | 0.28 |
| A | D | 0.37 |
| B | C | –0.05 |
| B | D | 0.33 |
| C | D | 0.05 |
| *Average* | | 0.22 |
| E | F | –0.05 |
| E | G | 0.00 |
| E | H | –0.02 |
| F | G | –0.16 |
| F | H | –0.15 |
| G | H | 0.01 |
| *Average* | | –0.06 |
| I | J | 0.08 |
| I | K | 0.12 |
| I | L | 0.00 |
| J | K | 0.24 |
| J | L | 0.18 |
| K | L | –0.18 |
| *Average* | | 0.07 |

General inspection of Tables 2 and 3 shows that the correlations between the separate tests are uniformly low, the second pattern (Tests E to H) being the most unsatisfactory. It would be interesting to pick out from each pattern the best tests if this could be done easily. This would properly be done by multiple correlation methods and the labor involved would not seem to be justified by the general promise of the coefficients reported. In pattern one, it would appear that Tests A and D would probably be found to be considerably superior to the other two. In pattern two, there is no promise that multiple correlation methods would reveal anything of significance. In pattern three, Tests J and K would appear the most promising.

## III Correlation of the Will Traits With General Intelligence

Table 4 shows the correlations obtained between the scores in the twelve separate will tests and the averages of two standard group tests. The latter are the Terman Group Test of Mental Ability and the Morgan Group Mental Test. The reliabilities of each of these tests is known to be well above 0.90 for ranges of talent comparable with those represented in the group considered here. The significance of the coefficients of correlations of Table 4, like those of all of the preceding tables, is obscured by the fact that the reliabilities of the separate tests are unknown and are probably very low in many cases.

### TABLE 4

CORRELATIONS OF THE TWELVE WILL TRAITS WITH MEASURES OF GENERAL INTELLIGENCE

| Test | r |
|---|---|
| A | 0.40 |
| B | 0.19 |
| C | 0.20 |
| D | 0.16 |
| E | −0.03 |
| F | 0.20 |
| G | 0.11 |
| H | 0.14 |
| I | 0.16 |
| J | 0.33 |
| K | 0.34 |
| L | −0.05 |

None of the above correlations is high enough to lend any support to the possible criticism of the Downey tests as being measures of general intelligence. The factor of intelligence undoubtedly does enter into some of the tests but not to an extent that would invalidate the scale on this score.

The inter-correlations between the tests have also been determined independently of the influence of intelligence by means of the usual methods of partial correlation. Table 5 shows such partial coefficients. The notations "$r_{AB.i}$", "$r_{AC.i}$", etc., read: The correlation between traits A and B which is independent of the influence of intelligence, and similarly, for the correlation between traits A and C, and the others.

## TABLE 5

THE CORRELATIONS BETWEEN THE SEPARATE TRAITS WHICH EXISTS INDEPENDENTLY OF THE FACTOR OF INTELLIGENCE (OBTAINED BY PARTIAL CORRELATIONS)

| | | | | | |
|---|---|---|---|---|---|
| $r_{AB.i}$ | 0.33 | $r_{EF.i}$ | −0.04 | $r_{IJ.i}$ | 0.03 |
| $r_{AC.i}$ | 0.22 | $r_{EG.i}$ | 0.00 | $r_{IK.i}$ | 0.07 |
| $r_{AD.i}$ | 0.34 | $r_{EH.i}$ | −0.02 | $r_{IL.i}$ | 0.01 |
| $r_{BC.i}$ | −0.09 | $r_{FG.i}$ | −0.17 | $r_{JK.i}$ | 0.14 |
| $r_{BD.i}$ | 0.31 | $r_{FH.i}$ | −0.16 | $r_{JL.i}$ | 0.21 |
| $r_{CD.i}$ | 0.02 | $r_{GH.i}$ | 0.00 | $r_{KL.i}$ | −0.18 |

By referring to Table 4 again it will be seen that there has been little change resulting from the use of partial correlations in eliminating the influence of the factor of intelligence. It may be concluded that whatever inter-correlation does exists between the various separate tests of the same patterns is not produced by the common factor of intelligence. The formula used in partial correlation is the usual one, following Yule's notations:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

## IV The Validity of the Downey Scale

The validation of any test of psychology capacities is at best a most difficult task. To this rule the Downey scale is probably no exception. For an account of the derivation of these tests the reader is referred to the manual for use with the Will-Temperament Scale[2].

In the earlier paper to which reference has several times been made, it was attempted to secure ratings from competent judges acquainted with the subjects on each of the twelve traits of the Downey scale. A similar procedure has been adopted in the present study with certain modifications. It was thought that the teachers of the high school pupils under study might be able to furnish a set of ratings on the twelve traits which would have

[2]Downey, June E.: "The Will-Profile." University of Wyoming Bulletin, Vol XVI, No. 4b, November, 1919.

value as a criterion against which to evaluate the tests proper. In adopting this plan, the authors were by no means ignorant of the many pitfalls in subjective judgments. As shown in certain of the published studies, it is possible, at times, with sufficient care, to obtain reliable subjective estimates in the sense of high self-correlation among the judges. However, high reliability, although admittedly highly desirable for such purposes, does not guarantee that a valid criterion has been built up. The question of the validity of such criteria is by far the more serious issue. The fact that ten persons rating a group of individuals for "initiative," "leadership," or "tact" agree with a reliability equal to 0.75 or higher means merely that they are rating the same thing or some other thing highly correlated with this thing, but it does not follow of necessity that the ratings correspond to the psychological functions which by hypothesis underlie the thing being rated. Teachers can furnish ratings on their pupils upon such traits as "persistence," "conscientiousness," "originality," "honesty," and a host of other traits with a fair amount of reliability, yet careful statistical analysis will reveal that little else is involved in these various separate trait ratings than the teachers' general estimates of the quality of school work being done by these pupils. In other words, almost as good results could be obtained by asking for estimates of school work and then using these, in turn, as estimates of "persistence," "originality," "conscientiousness," etc. The estimates, then, are clearly not valid except as these traits are genuinely correlated with the quality of school work and even then it is a very indirect type of measurement.

On the other hand what conclusions can be drawn from the situation of judges not being able to agree upon the presence or absence of quantitative statement of psychological traits in a group of individuals known to the judges? Does failure to show self-correlation in chance halves of these sets of estimates carry with it any important implication? Does such a negative result with respect to a psychological trait prove that the trait in question has no true psychological basis? No universal answer can be given to these queries. In the case of many reputed psychological rubrics it probably does mean that there is no real evidence for assuming the reality of such traits as true mental functions. In other cases if not in most the answer is left in an undeterminate status, i.e., the experimental results are entirely non-committal in the sense that they furnish no evidence, positive or negative, upon the issue of the existence or non-existence of the trait in question. Even at this it might for all practical purposes be argued that unless competent persons are in fair agreement about the presence

or absence of reputed psychological characteristics there is no direct way of identifying such traits for purposes of study.

With these limitations of the method of estimates in mind we can return to the discussion of the immediate problem. In the attempt to build up a criterion for purposes of correlation of the trait scores with a second measure of the same traits the high school teachers of the pupils in our present study were asked simply to nominate the ten pupils in their classes who possessed the "strongest will-power" and the ten who had the "weakest will-power." No attempt at further definition of these directions was made. The reactions expected were those of the man on the street. The only outcome which was expected was that the pooled nominations might reveal two groups widely spaced upon a scale of "will-power" in the common meaning of that term. When the nomination blanks were received it was found that certain pupils were invariably reported in the same category. A larger number were about as often placed in one group as the other and a still larger number were never once nominated for either classification. There was, then, evidence of marked agreement in many cases, the total number of complete agreements being roughly 20 per cent of the total of one hundred forty-six pupils in the high school under study. Twenty-eight pupils involving equal numbers of the sexes and equal numbers in each of the two categories (i.e., "strong-willed" and "weak-willed") were selected for further study. The agreement was perfect or practically perfect in these cases in the sense that either three or four of the four teachers acquainted with every pupil placed him in the same category and no teacher gave a dissenting vote in any of these cases. The reliability coefficient for this final group was not determined but its general nature is assuredly high by virtue of the method of forming the group. Instructions were then sent to every teacher who had nominated any of the pupils in the final group to rate each pupil upon each of the twelve traits of the Downey scale. The trait descriptions were taken verbatim from the author's manual for use with these tests.

The final group therefore comprised (to the best knowledge of all of the teachers of the school) the fourteen pupils of the school showing the most marked volitional characteristics and an equal number most lacking in these same characteristics. As well as could be done by estimation we had segregated the highest and lowest ten per cents of the group as a whole. Table 6 shows the correlations found between the pooled estimates of the teachers and the test scores in each of the twelve traits. Inspection of these coefficients of correlation does not reveal very definite evidence of the ability of teachers to identify the various traits tested

by the Downey scale. Whether this situation can be interpreted as meaning that the Downey scale is of doubtful validity is very questionable. A saner conclusion would be that teachers are unable to identify the traits of the Downey scale in their pupils to any considerable extent (traits J and K, and possibly traits A and C being apparent exceptions). It is probably fair to assume, however, that the Downey tests cannot be validated by the method of subjective estimates with any high amount of assurance. The correlations between estimates and test scores are in general of about the same magnitudes as those reported in the earlier study (where the range was from –0.33 to 0.51 for the pooled estimates of three or four college professors; the average being 0.07). The average estimate-test correlation is 0.15 for Table 6.

## TABLE 6

CORRELATIONS BETWEEN ESTIMATES AND TEST SCORES (N=28)

| Trait | r | Trait | r | Trait | r |
|---|---|---|---|---|---|
| A | 0.37 | E | 0.19 | I | –0.09 |
| B | 0.02 | F | –0.03 | J | 0.51 |
| C | 0.30 | G | –0.04 | K | 0.53 |
| D | –0.29 | H | 0.17 | L | 0.17 |

The probable errors of the coefficients of correlation in Table 6 can be found in Table 7 to a sufficient accuracy for our purposes. The same is true of the correlation coefficients given in Tables 1 to 5. Except in Table 6, the numbers of cases have invariably been one hundred and forty-six.

## TABLE 7

A REFERENCE TABLE OF PROBABLE ERRORS OF PEARSON COEFFICIENTS OF CORRELATION FOR VALUES OF r APPEARING IN THE FOREGOING TABLES

| r | Probable Erors of r | |
|---|---|---|
| | N=146 | N=28 |
| 0.00 | .056 | .127 |
| 0.05 | .056 | .127 |
| 0.10 | .055 | .126 |
| 0.15 | .055 | .125 |
| 0.20 | .054 | .122 |
| 0.25 | .052 | .120 |
| 0.30 | .051 | .116 |
| 0.35 | .049 | .112 |
| 0.40 | .047 | .107 |
| 0.45 | .045 | .102 |
| 0.50 | .042 | .096 |
| 0.55 | .038 | .089 |

## CONCLUSIONS.

The following points are suggested as summarizing the more important findings presented:

(1) The reliability of the *Downey Group Will-Temperament Test* is probably not higher than 0.40 as an estimate of the correlation which would be found between two similar forms of the tests were a second form available. It is no doubt true that certain of the patterns are more reliable than others, a fact which militates against the reliability of the scale as a whole.

(2) The validity of certain of the individual tests is doubtful. The inter-correlations in patterns II and III are low, pattern I being the most satisfactory. The same situation holds for the correlations of the individual tests with the composites of the remaining tests of the same pattern.

(3) The lack of inter-correlation between tests of the same pattern in the case of pattern II, and to a less degree in the other patterns, may be due to either: (a) unreliability of the individual tests (not determined in this study), or (b) faulty location of certain tests. In the latter case further study might suggest changes from one pattern into another.

(4) There was found no reason to believe that the Downey tests measure merely intelligence. The inter-correlations, where they are significally high, are not due to the common factor of intelligence.

(5) The attempt to validate the separate tests of the scale by means of teachers' judgments yielded ambiguous results. It would appear that strict validation of the Downey tests by the method of estimates would be a very difficult task. We can conclude that, with a few exceptions, high school teachers cannot identify the psychological traits measured by the Downey tests by study of the characterizations of those traits which have been published by the author of the tests. It should be pointed out that the estimates used in this study cannot be assumed to be valid, although their general reliability is doubtless satisfactorily high for purposes of a rough criterion. The question of the validation of the Downey tests can better be dismissed by refering the reader to Dr. Downey's manual, previously cited, for a discussion of this point.

(6) In general there are a considerable number of tests in the *Downey Will-Temperament Test* that show promise of lending themselves to further study and development. Duplication of such tests and probably a number of substitutions will lead to a more valid instrument for the measurement of volitional

traits. The urgent need for such measurements demands that the pioneer efforts of Dr. Downey be rewarded with the recognition that much promising ground had been broken by her patient investigations and that further progress has been made possible in this direction.

# THE FRESHMAN: THORNDIKE COLLEGE ENTRANCE TESTS, FIRST SEMESTER GRADES, BINET TESTS

By W. T. Root,
University of Pittsburgh.

*INTRODUCTION.* At the beginning of the academic year, 1921-22, the College of the University of Pittsburgh* gave the Thorndike Intelligence Examination for High School Graduates to a Freshman class of 569. The tests given were: Part I, Form I (designated as T1); Part I, Form M (T2); Part II, Form D (T3); Part III, Form D (T4). The total, that is the sum of the four tests, is designated T5. The test was given without intermission other than the short intervals provided for in the instructions.

Results. The following averages were secured:

| Test | Average |
|---|---|
| T1 | 105.119 |
| T2 | 105.276 |
| T3 | 101.504 |
| T4 | 32.97 |
| T5 | 341.61 |

Final score (T5 divided by 5) 68.38
Median score, 70
Range 21 to 110

*Correlations between different forms.* The forms were correlated with the total singly and in combination:

| Test | r (Pearson) | P. E. |
|---|---|---|
| T5 with T1 | .848 | .007 |
| T5 with T2 | .829 | .009 |
| T5 with T3 | .904 | .005 |
| T5 with T4 | .754 | .01 |
| T5 with T1 plus T2 | .914 | .004 |
| T5 with T1 plus T3 | .960 | .002 |

*The Department of Psychology gave and scored the tests. The writer is indebted to the graduate students in psychology for many weary hours of statistical calculation. Special acknowledgment is due Professor Florence M. Teagarden for transmuting grades into sigma values and for giving the Binet Tests to the Freshmen girls. Mrs. Ida Nyce Root is to be thanked for many days spent in the vexatious task of statistical rechecking and of ferreting out errors.

*The high correlation of T1 plus T3 with T5 suggests very strongly that T2 and T4 be eliminated in future testing. This would mean a great economy in the time required to give the test, in the time required to score, and in the cost of test materials.*

*Correlation of T5 with the academic grades for the first semester.* In attempting to summate the academic grades the writer met with the usual difficulties. It seemed best after all factors were considered to resort to the tedious method of transmuting all grades to sigma values thus correcting for the great variance in distribution of A's, B's, C's, D's, and F's. The sigma value was weighted according to the number of units of credit the course in question carried. The weighted sigma values were added and divided by the number of units of credit carried by the particular Freshman in question. This procedure overcomes the injustice usually experienced by students carrying "stiff" courses (i.e., courses with a high percentage of D's and F's and a low percentage of A's and B's). The sigma values of A, B, C, D, and F were determined for each subject taught to Freshmen. D is the lowest passing grade while F is failure.

The correlations between the academic grades and the Thorndike forms are as follows:

| Test | | r | P.E. |
|---|---|---|---|
| T1 | with grades | .42 | .02 |
| T2 | with grades | .33 | .03 |
| T3 | with grades | .36 | .03 |
| T4 | with grades | .415 | .02 |
| T5 | with grades | .51 | .02 |

The Thorndike total was now correlated with each college subject taken by the Freshmen, debarring a few in which the number of beginning students was too small to admit of correlation. (Tabulation I)

Academic College Subjects: Grade Percentages, Sigma Values and Correlations With Thorndike College Entrance Tests.

| College Subjects | Units of Credit | GRADES | | | | | | | | | | | | Correlations between Academic Grades and T. College Entrance Tests | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | | | B | | | C | | | D | | | F | | | Number of Cases | r | P. E. of r |
| | | Percentage of | σ Value | Weighted σ Value | Percentage of | σ Value | Weighted σ Value | Percentage of | σ Value | Weighted σ Value | Percentage of | σ Value | Weighted σ Value | Percentage of | σ Value | Weighted σ Value | | | |
| Biology | 1 | 4 | 12.16 | 4.54 | 18.12 | 31.08 | 3.59 | 14.36 | 28.37 | 2.81 | 11.24 | 20.27 | 2.10 | 8.40 | 8.10 | 1.27 | 5.08 | 73 | .495 | .06 |
| Biology | 51 | 4 | 6.25 | 4.84 | 19.36 | 21.25 | 3.95 | 16.80 | 30.00 | 3.19 | 12.76 | 27.50 | 2.44 | 9.76 | 15.00 | 1.57 | 6.28 | 151 | .525 | .05 |
| Chemistry | 1 | 4 | 4.00 | 5.03 | 20.12 | 32.00 | 3.80 | 15.02 | 26.00 | 3.00 | 12.00 | 16.00 | 2.45 | 9.80 | 21.00 | 1.75 | 7.00 | 258 | .434 | .03 |
| English | 1 | 3 | 6.00 | 4.86 | 14.58 | 27.00 | 3.86 | 11.58 | 50.00 | 2.81 | 8.43 | 15.00 | 1.70 | 5.10 | 2.00 | .73 | 2.19 | 495 | .358 | .08 |
| French | 1 | 3 | 13.00 | 4.50 | 13.50 | 23.00 | 3.69 | 11.07 | 31.00 | 2.97 | 8.91 | 21.00 | 2.25 | 6.75 | 12.00 | 1.46 | 4.38 | 72 | .396 | .07 |
| French | 3 | 3 | 13.00 | 4.50 | 13.50 | 32.00 | 3.55 | 10.65 | 37.00 | 2.66 | 7.98 | 14.00 | 1.78 | 5.34 | 4.00 | .97 | 2.91 | 131 | .453 | .05 |
| French | 7 | 3 | 21.00 | 4.19 | 12.57 | 43.00 | 3.16 | 9.48 | 14.00 | 2.42 | 7.26 | 14.00 | 1.93 | 5.79 | 7.00 | 1.21 | 3.63 | | | |
| German | 1 | 3 | 10.00 | 4.63 | 13.89 | 25.00 | 3.75 | 11.25 | 30.00 | 3.00 | 9.00 | 20.00 | 2.33 | 6.99 | 15.00 | 1.57 | 4.71 | 60 | .503 | .06 |
| Graphics | 2 | 2 | 3.00 | 5.14 | 10.28 | 8.00 | 4.47 | 8.94 | 40.00 | 3.50 | 7.00 | 43.00 | 2.47 | 4.94 | 6.00 | 1.14 | 2.28 | 109 | .348 | .06 |
| History | 1 | 3 | 11.00 | 4.59 | 13.77 | 25.00 | 3.72 | 11.16 | 47.00 | 2.76 | 8.28 | 16.00 | 1.67 | 5.01 | 1.00 | .51 | 1.53 | 122 | .483 | .05 |
| History | 7 | 3 | 9.00 | 4.68 | 14.04 | 22.00 | 3.84 | 11.52 | 44.00 | 2.93 | 8.79 | 14.00 | 2.09 | 6.27 | 11.00 | 1.41 | 4.23 | 48 | .429 | .08 |
| History of Religion | | 2 | 9.00 | 4.68 | 9.36 | 32.00 | 3.67 | 7.34 | 41.00 | 2.72 | 5.44 | 18.00 | 1.67 | 3.34 | 0.00 | 0.00 | 0.00 | | | |
| Human Prog.[1] | | 2 | 4.00 | 5.03 | 10.06 | 29.00 | 3.89 | 7.78 | 45.00 | 2.87 | 5.74 | 19.00 | 1.86 | 3.72 | 3.00 | .86 | 1.72 | 527 | .689 | .01 |
| Latin | 5 | 3 | 21.00 | 4.25 | 12.75 | 42.00 | 3.20 | 9.60 | 25.00 | 2.31 | 6.93 | 8.00 | 1.60 | 4.80 | 4.00 | .97 | 2.91 | | | |
| Math. | A | 3 | 5.71 | 4.88 | 14.64 | 22.86 | 3.94 | 11.82 | 22.86 | 3.25 | 9.75 | 31.43 | 2.56 | 7.68 | 17.14 | 1.64 | 4.92 | 33 | .578 | .08 |
| Math. | 1 | 3 | 25.49 | 4.13 | 12.39 | 13.72 | 3.45 | 10.35 | 21.57 | 3.00 | 9.00 | 25.49 | 2.38 | 7.14 | 13.73 | 1.52 | 4.56 | 52 | .512 | .07 |
| Math. | 2 | 3 | 12.50 | 4.52 | 13.56 | 12.50 | 3.88 | 11.64 | 33.33 | 3.21 | 9.63 | 29.16 | 2.39 | 7.17 | 12.50 | 1.48 | 4.44 | 23 | .614 | .09 |
| Math. | 21 | 4 | 6.31 | 4.84 | 19.36 | 15.32 | 4.08 | 16.32 | 25.22 | 3.40 | 14.60 | 32.43 | 2.67 | 10.68 | 20.72 | 1.75 | 7.00 | 110 | .390 | .06 |
| Physics | 1 | 4 | 10.00 | 4.63 | 18.52 | 17.00 | 3.89 | 15.56 | 31.00 | 3.19 | 12.76 | 21.00 | 2.52 | 10.08 | 21.00 | 1.75 | 7.00 | 134 | .496 | .04 |
| Psychology | 3 | 3 | 18.00 | 4.33 | 12.99 | 9.00 | 3.75 | 11.25 | 18.00 | 3.35 | 10.05 | 46.00 | 2.54 | 7.62 | 9.00 | 1.32 | 3.96 | | | |
| Pub. Speaking | 1 | 2 | 10.61 | 4.06 | 8.12 | 42.47 | 2.95 | 5.92 | 42.47 | 1.86 | 3.72 | 3.53 | .65 | 1.30 | .87 | 19.00 | .38 | 108 | .458 | .05 |
| Spanish | 1 | 3 | 12.00 | 4.54 | 13.62 | 19.00 | 3.78 | 11.34 | 32.00 | 3.07 | 9.21 | 20.00 | 2.39 | 7.17 | 17.00 | 1.64 | 4.92 | 71 | .472 | .06 |
| Spanish | 3 | 3 | 19.00 | 4.30 | 12.90 | 22.00 | 3.52 | 10.56 | 25.00 | 2.92 | 8.76 | 16.00 | 2.36 | 7.08 | 18.00 | 1.67 | 5.01 | 14 | .666 | .08 |

[1]Human Progress is a required Freshman Course similar to the course on Contemporary Civilization offered at Columbia University.

The correlations are moderately high and considerable significance as N is large in the majority of cases and the group is homogeneous. The correlation of the Thorndike test with Human Progress is conspicuously high (r. .689) with a very small probable error (P. E. .01). This course deals with a large number of the arts and sciences in a semi-philosophical way. The classes are large and a premium is placed on general information and linguistic ability. English 1, Graphics 2, Mathematics 21 and French 1 correlated less highly than the other courses with the Thorndike total. English may be a matter of home conditions, foreign parentage or drill and follow up work through four years of high school. The students enter Graphics and Mathematics with very different preparations; French may be very easy for the student with a fair knowledge of Latin or one of the Latin languages. If Italian is spoken in the home it may prove to be a hinderance in an English course and a boon in a French course. One may speculate indefinitely as to the general causes for correlation or non-correlation.

*The prognostic value of the Thorndike College Entrance Tests.* It is possible to have reasonably high correlations without much justification for applying the findings to an individual case. Distinct sets of causes may be at work: thus, we may have nearly perfect correlation with a portion of the cases, a moderately positive trend for a large median group, with a significant group showing a negative correlation. Again, we may have marked negative or positive correlation in certain regions of the data with an indifferent correspondence in other regions. It becomes pertinent, then, to ask for an *analysis* of the tendency to correlation or rank-correspondence as well as to compute the coefficient of correlation.

The chief reason for administering the tests now under consideration to Freshmen is to predict their capabilities in connection with academic college requirements. It therefore seemed desirable to rate the Freshmen in some form of grade classification, arranged in rank-order, and to compare this ranking with the Thorndike rank-orders. The conventional distribution of grades over a $\pm 3\alpha$ normal distribution curve is, perhaps, as good as any. The data can then be tabulated as follows (A, B, C, D and F referring to grades and $\alpha$, $\beta\gamma$ and $\epsilon$ referring to precisely the same percental divisions in the descending rank-order arrangement of the Thorndike totals):

| | A or $\alpha$ | B or $\beta$ | C or $\gamma$ | D or $\delta$ | F or $\epsilon$ |
|---|---|---|---|---|---|
| % ..... | 3.5 | 24 | 45 | 24 | 3.5 |
| No. cases..... | 20 | 137 | 255 | 137 | 20 |

The college grades are expressed as sigma values and each Freshman's rank-order is determined by finding his average sigma (the sum of his weighted sigmas divided by the number of units carried). These average sigma values arranged in descending rank-order are then ready for comparison with the corresponding rank-orders in the Thorndike tests. If the Thorndike test is prognostic of academic ability then the 20 students composing the A group should have a rank-correspondence such as to place the majority of them in the $\alpha$ group (the 20 ranking highest in the Thorndike group). Tabulation II presents the rank-orders for college grades and Thorndike scores for comparison and analysis of the rank-correspondence in different percental regions of the rank-order series. The question is: Does the Thorndike Test performance anticipate the quality or approximate quality of college work a particular Freshman will do?

TABULATION II.

| | A | B | C | D | F | TOTALS |
|---|---|---|---|---|---|---|
| A | 4 1 | 9 5 | 7 9 | 0 5 | 0 1 | 20 |
| B | 7 5 | 60 33 | 54 61 | 11 33 | 5 5 | 137 |
| Γ | 8 9 | 62 61 | 115 115 | 65 81 | 5 5 | 255 |
| Δ | 1 5 | 6 33 | 70 61 | 53 33 | 7 5 | 137 |
| E | 0 1 | 0 5 | 9 9 | 8 5 | 3 1 | 20 |
| TOTALS | 20 | 137 | 255 | 137 | 20 | 569 |

The following observations may be made directly from the tabulated data:

1. No. $\alpha$'s are ranked D or F in academic standing.
2. No. $\epsilon$'s attain A or B standing.
3. But of those making A standing, only 1/5 are $\alpha$'s and the largest number are drawn from the $\gamma$ group.
4. Had superior scholarship been predicted by selecting the $\alpha$ group we would actually have secured 4 A students along with 9 B students and 7 C students.
5. 80 of the 157 students comprising the A and B groups are in the $\alpha$ and $\beta$ groups.

Seventy-one of the 157 comprising the D and F groups are in the $\delta$ and $\epsilon$ groups.

That is, the upper 27.5% of the Thorndike rank-orders would have been somewhat selective as regards academic attainment. Had the $\epsilon$ group been taken as a basis of entrance exclusion, it would have resulted in the excluding of 8 D students and 9 C students out of a total of 20 $\epsilon$ students.

6. Had the college wished to maintain very high standards and had excluded $\delta$ and $\epsilon$ standings, it would have excluded 79 C students, 6 B students, and 1 A student out of a total of 157 Freshmen. The Thorndike ranking would have correctly placed 71 (that is $\delta$-$\epsilon$'s who are in corresponding academic rank-orders, namely the D-F groups), while random choice alone would have correctly placed 44 of these.

7. On the other hand, if the $\alpha$-$\beta$-$\gamma$ groups had been used as a basis of initial retention in college, 86 D and F students would have been admitted into the Freshman class against the 71 D and F students excluded in the $\delta$-$\epsilon$'s exclusion. Chance alone would have given us 114 D and F grade Freshmen, out of a total of 410 $\alpha$-$\beta$-$\gamma$'s.

Of the 569 students chance alone would have given 157 DF's. Had the $\delta$-$\epsilon$ groups been arbitrarily excluded (and the $\alpha$-$\beta$-$\gamma$ groups retained), 71 D-F students would have been excluded and 86 D-F included. But 86 A-B-C students would have been thus excluded. This would give a total of 176 improper inclusions and improper exclusions or 19 more than the improper inclusions chance alone would have given.

8. There are five comparable groups, $\alpha$-A, $\beta$-B, $\gamma$-C, $\delta$-D, and $\epsilon$-F. Of these groups, the $\gamma$-C is the only one that is largest for the column in question. In this case it is totally without prognostic significance as chance alone would have placed 115 C's in the $\gamma$ group. It should furthermore be noted that the distribution of C students through $\alpha$, $\beta$, $\delta$, $\gamma$ and $\epsilon$ is about the distribution

chance would have given. And likewise the row has a close-to-chance distribution through A, B, C, D and F. Hence for a student to be in the $\gamma$ group argues nothing as to his academic rank-order; and vice versa, C grades run the gamut of a near-chance distribution curve from $\alpha$ to $\epsilon$.

Deducting from each of the five comparable groups (A-$\alpha$, B-$\beta$, etc.) the number of cases that would fall in these groups by chance alone, gives the number of cases precisely located in a corresponding rank-order. We have:

Of 20 $\alpha$'s, 3 A's or 15% in precise corresponding rank-order.
Of 137 $\beta$'s, 27 B's or 19% in precise corresponding rank-order.
Of 255 $\gamma$'s, none.
Of 137 $\delta$'s, 20 or 15% in precise corresponding rank-order.
Of 20 $\epsilon$'s, 2 or 10% in precise corresponding rank-order.

9. Compiling $\beta$, $\gamma$ and $\delta$ we have 529 cases, comprising 128 B's, 246 C's, and 129 D's. Thus the Thorndike tests would have correctly placed 12 more subjects than chance alone or about 2 per cent of the total number of students in the combined $\beta$, $\gamma$ and $\delta$ groups.

TABULATION III.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| A | Scores over 95 | 9 | 2 | 0 | 3 | 5 | 1 | 0 | 0 |
| B | Scores 85 to 95 | 55 | 10 | 2 | 17 | 24 | 7 | 1 | 4 |
| C | Scores 70 to 85 | 203 | 36 | 6 | 25 | 92 | 45 | 23 | 12 |
| D | Scores 60 to 70 | 143 | 25 | 0 | 8 | 53 | 37 | 27 | 18 |
| E | Scores 50 to 60 | 96 | 17 | 1 | 2 | 18 | 27 | 34 | 14 |
| F | Scores under 50 | 63 | 11 | 0 | 0 | 11 | 26 | 11 | 15 |

AN ANALYSIS OF THORNDIKE'S GROUPINGS

Thorndike attempts to prophecy on a basis of the absolute scores received. He makes the following classification:

"A boy scoring over 95 is worth admitting in almost entire disregard of technical deficiencies.

A boy scoring 85 to 95 has intellect enough to do collegiate and professional work with distinction.

A boy scoring 70 to 85 has intellect enough to do the work to obtain a college degree.

A boy scoring 60 to 70 may be admitted if he is sufficiently in earnest and otherwise desirable.

A boy scoring 50 to 60 should be admitted only if he is of extraordinary zeal or has suffered very great educational handicaps.

A boy scoring under 50 should not be admitted."

The accompanying tabulation (III) follows Thorndike's grouping. The academic grade rank-orders have been arbitrarily classified to agree in rank correspondence with the Thorndike classification.

Conclusions:

1. None of the Thorndike A group falls in the corresponding academic group a (the 9 highest academic rank-orders); however, none falls in the lowest academic groups, e and f.

2. Likewise none of the F group falls in the academic a or b groups.

3. The students composing the a group came chiefly from the C group.

4. Taking Thorndike's C group as the dividing line for inclusion in or exclusion from college entrance, one would include 174 correctly but also 93 incorrectly, (students making d, f, and e grades). On the other hand, the elimination of the D, E, and F students would have excluded 93 students who are doing work in the a, b and c groups. Thus there would be 186 improper inclusions and exclusions in using Thorndike's A, B and C groups as an inclusion basis against 209 proper exclusions. Out of 569 students 302 would be excluded. It must be recalled that in excluding such a large percentage of the students seeking admission, chance alone would have correctly excluded 161.

5. Excluding E's and F's would retain 85 of the upper Thorndike groups doing no better work than E's and F's, and would exclude 85 E's and F's doing work on a par with those admitted. Granting that e and f grade work warrants exclusion, we have 74 proper exclusions, and 170 improper inclusions and unjust exclusions.

6. Let us take Thorndike's lowest group, F, (scores below 50), which Thorndike categorically asserts should be excluded. Of the 63 comprising this group 48 do work on a par with many in the retained groups, while 48 in the retained groups are doing the same quality of academic work as those excluded. Thus we have 96 improper exclusions and inclusions, and 15 proper exclusions.

7. Tabulations II and III lead to the same general conclusions. *While the correlation between grades and Thorndike scores is reasonably high, an inspection of the differences in rank-order*

*correspondence indicates that great injustice would be done, both in inclusions and exclusions, were the entrance tests used as a basis of college admission. In actual practice the lines of demarcation between groups are too illy defined to warrant such drastic action as off-hand exclusion from college. However, the trend of correlation is sufficiently strong to far surpass any impressionistic estimate and the tests should constitute an excellent basis for the initial analysis of cases, and for judicious use in attempting to size up Freshmen failures. Whether the mass test is more truly prophetic of final college success than the combined estimates of dozens of instructors after seventeen weeks contact is a speculative matter which can be decided only by follow-up work for a number of years.*

### EXTREME UPPER AND LOWER RANK-ORDERS IN COLLEGE ENTRANCE TESTS COMPARED AS TO PERFORMANCE IN BINET TESTS

Forty of the upper rank-orders of the Thorndike Test (1to 58 inclusive) and also 40 of the lower rank-orders (479 to 569 inclusive)[2] were given the Binet Tests.[3] The accompanying tabulation (III) shows clearly that the correspondence between the two tests is extremely close, quite in contrast to the vague correspondence between the Thorndike and the academic standing.

[2]Inability to arrange a time suitable for both the student and the examiner giving the Binet Test, made it necessary to go below rank 40 and above rank 529 in the lower group. One of the lower rank-orders was eliminated. The subject, an intelligent Italian just learning the language, refused to try the Binet Test after doing the vocabulary.

[3]The student appointments for the Binet Tests were arranged through the co-operation of Professor B. F. Ashe, Freshman Advisor.

## TABULATION IV

Binet Tests Used as a Basis for Comparing the Extreme Upper and Lower Rank Orders of Thorndike College Entrance Tests

| | | 40 highest Rank Orders | 39 lowest Rank Orders |
|---|---|---|---|
| Chronological age range | | 22-8 to 16-5 | 27 to 18-2 |
| Chronological age median | | 20-1 | 18-5 |
| Mental age range | | 19-6 to 17-2 | 18 to 12-1 |
| Mental age median | | 19 | 15-5 |
| Vocabulary range | | 16,560 to 12,780 | 12,600 to 6,840 |
| Vocabulary median | | 14,940 | 10,260 |
| Number failing | | | |
| XVIII | Superior adult vocabulary | 5 | 39 |
| | Binet's paper-cutting test | | 25 |
| | Repeats 8 digits | 13 | 24 |
| | Repeats thought of passage | | 26 |
| | Repeats 7 digits backward | 12 | 31 |
| | Ingenuity test | 8 | 31 |
| XVI | Average adult vocabulary | | 30 |
| | Interpretation of fables | | 15 |
| | Difference between abstract words | | 11 |
| | Problem of the enclosed boxes. | 5 | 22 |
| | Repeats 6 digits backwards | 1 | 17 |
| | Code | | 23 |
| XIV | Vocabulary 50 words | | 8 |
| | Induction test | | 7 |
| | President and king | | 8 |
| | Problems of fact | | 10 |
| | Arithmetical reasoning | | 9 |
| | Clock | | 8 |
| XII | Vocabulary 40 words | | 1 |
| | Abstract words | | |
| | Ball and field | | 4 |
| | Dissected sentences | | 3 |
| | Fables | | 1 |
| | Repeats 5 digits backwards | | 3 |
| | Pictures interpretation | | 1 |
| | Gives similarities | | 1 |
| X | Vocabulary 30 words | | |
| | Absurdities | | 1 |
| | 60 words | | 2 |
| Total No. of tests failed | | 44 | 361 |

The lower group is made up of older students and it may be added has more of foreign birth and foreign family. It is a matter of general observation that older persons show less associative plasticity and speed.

Concerning I. Q. and mental age:

8 of the lower group had I. Q.'s 100 or more.
31 of the lower group had I. Q.'s less than 100.
3 of the lower group had mental ages of 12(11-7 to 12-6)
5 of the lower group had mental ages of 13(12-7 to 13-6)
4 of the lower group had mental ages of 14(13-7 to 14-6)
8 of the lower group had mental ages of 15(14-7 to 15-6)
12 of the lower group had mental ages of 16(15-7 to 16-6)
4 of the lower group had mental ages of 17(16-7 to 17-6)
2 of the lower group had mental ages of 18(17-7 to 18-6)

In the upper group the lowest I. Q. was 104. Nineteen failed in the high adult digits tests; 7 failed in both, 14 passed all of the tests; 13 failed 1 test; 11 failed 2 tests; 1 failed 3 tests; and 1 failed 5 tests.

The fact that 25 of the 44 failures in the upper group were digit tests indicates that the digit tests (counting an entire year of mental age) are entirely out of proportion to their importance in determining mental age. A careful study of the individual cases gives little evidence that immediate verbal auditory memory span for digits is of any marked value in determining superior intelligence. It is a matter in which there is marked individual difference in adults with slight reason for believing that there is a corresponding intellectual difference.

*Reading and Vocabulary.* The most striking contrasts between the groups is in the vocabularies and reading habits. The highest vocabulary in the low group is less than the poorest vocabulary of the high group.

An attempt was made to get at the reading habits of the groups under consideration. A rough reading scale was improvised, scoring as follows:

1. Readers of critical and difficult literature.
2. Superior readers, read classical and good current literature.
3. Average readers.
4. Readers of newspapers and sport news.
5. Readers who do almost no voluntary reading being nearly devoid of a reading habit.

Tabulating for the two groups we have:

| Reading Rank | Number in High Group | Number in Low Group |
|---|---|---|
| 1 | 6 | 1 |
| 2 | 12 | 2 |
| 3 | 15 | 7 |
| 4 | 7 | 14 |
| 5 | 0 | 15 |

The contrast here is very marked indicating decided lack of reading habits in the lower group.

## TABULATION V.

Cases With Marked Inconsistency in Performance in Upper and Lower Groups

| Thorndike Test Rank | Academic Rank | Reading Rank | Voc. Score | Mental Age Binet | REMARKS |
|---|---|---|---|---|---|
| 3 | 386½ | 3 | 91 | 19 | Out of school 3 yrs. Trig. & Physics difficult |
| 5 | 132½ | 2 | 92 | 19 | |
| 6 | 137 | 3 | 92 | 19 | |
| 8 | 188½ | 1 | 86 | 18½ | Out of school 3 yrs. Indifferent to grades |
| 10 | 197½ | 2 | 83 | 19½ | Exemption grades in H. S. |
| 13.5 | 188½ | 3 | 81.5 | 19 | |
| 20 | 297 | 1 | 87 | 19½ | |
| 27½ | 223½ | 2 | 79 | 19 | |
| 30 | 197 | 1 | 90½ | 19 | |
| 35½ | 554 | 2 | 88 | 19 | Absent one month in first semester |
| 35½ | 301½ | 4 | 84 | 19½ | |
| 40½ | 542 | 2 | 81 | 18½ | |
| 48½ | 350½ | 3 | 73 | 18½ | |
| 58½ | 258 | 4 | 89.5 | 17-2 | |
| 522 | 283 | 3 | 58 | 13-6 | Italian. Foreign parent. |
| 524 | 71 | 2 | 56 | 16-5 | Jewish. Soho district. |
| 526 | 172 | 2 | 67 | 15-8 | Jewish. Foreign parent. |
| 506½ | 290½ | 3 | 70 | 17-1 | Attended sixteen different schools. |
| 538½ | 249 | 4 | 62 | 16 | |
| 545½ | 267½ | 5 | 40 | 14-9 | Italian. Foreign home. |
| 553½ | 264 | 5 | 63 | 13-11 | Jewish. |
| 554 | 208½ | 4 | 70 | 17-7 | Jewish. |
| 557 | 157½ | 5 | 48 | 12-2 | Negro. |
| 561 | 206 | 4 | 68 | 18 | Commutes four hours daily. |

*Analysis of cases showing marked contrasts in performance.* It would have been highly desirable to have studied the social conditions surrounding these Freshmen composing the upper and lower groups. This was impossible, due both to lack of time and to the diffidence on the part of students concerning their social and economic status. Nevertheless, after personal interviews with the students in question, it is possible to throw some light on the marked inconsistencies in test performance and academic rank, presented in the accompanying Tabulation V.

## CONCLUSIONS

1. A high Thorndike rank is accompanied by a high Binet score and large vocabulary; and vice versa, a low Thorndike score by low mental age and small vocabulary.

2. But high vocabularies are not necessarily found with a superior reading ranking or vice versa.

3. Of the 40 in the upper Thorndike group, 14 are distinctly below any corresponding academic rank-order, while 7 fall below the median rank-order, 285, of the entire Freshman class.

4. In the lower group just the opposite is found; with low mental age, low vocabulary and low entrance examination score is found a relatively high academic rank-order. Of the 10 cases, 9 have academic ranks above the median rank-order of the Thorndike scores.

5. The cases showing marked inconsistencies constitute nearly 1-3 of the total number of the 79 cases studied intensively.

6. It should also be noted that in spite of exceedingly low vocabulary and mental age, cases 522, 553½, and 557 make more than median grades.

7. In the lower group, 6 are either Italian or Jewish. In all of these cases, and many more not presented here, the parents speak a foreign language. Is it not to be expected that an English vocabulary and tests resting on the niceties of English idiom will suffer?

8. Absence during the semester and a several-year gap between high school and college seem to be the most obvious and plausible explanations in certain cases of relative academic inferiority in the upper group.

9. The negro, case 557, unlike the Jewish and Italian cases, appears very dull. He is carrying, however, a very heavy program including mathematics, physics and chemistry in which he is making more than median grades. An examination of the percentage of A's, B's, C's, D's and F's given in these courses indicates that they are not "pipe courses," (see Tabulation I). In the opinion of the writer the explanation is this: The student has

had an excellent high school training with a careful drilling in technical courses that dovetail into the courses taken in college. It is doubtful if the chap will long maintain an academic ranking of 157½ against a Thorndike rank of 557 and a Binet mental age of 12-2.

10. The personal interviews elicited the following, at least plausible, explanations for the marked rank-order difference between academic grades and the Thorndike tests:

(a) Some students are capable but indifferent to grades, tending to follow an interest rather than make punctilious preparation of daily tasks. Case 8 is a good example. In the opinion of the writer, the Thorndike test places him more nearly correctly. The chap has been out of school three years between high school and college, which is perhaps a contributing factor, but the chief difficulty is that he has very positive ideas concerning academic values and refuses to fulfill, what he considers to be, petty classroom demands. He has already been successful in the work-a-day world.

(b) Outside work, undoubtedly, in many cases makes for poorer grades.

(c) Commuting tends, also, to produce a poorer grade of work than otherwise.

(d) The character of the high school work plays a large part in the character of the Freshman college work. For example, one high school, gives to its seniors a course on Current Literature. All of these students show a relatively greater general information than the average high school graduate. Another high school apparently gives unusual work in science and mathematics. One of the students from the lower Thorndike group, with good grades, stated that the physics and mathematics he was taking were almost entirely review. Again, a student changing from a literary or a scientific course in high school to a scientific or literary in college is greatly handicapped over the one who has had a constant vocational objective. All considered, students who have established good habits in mathematics, science, laboratory and English composition, regardless of their native endowment, have an advantage over those who have not.

(e) Unusual zeal for learning and persistence mark the attitude of many of the foreigners, especially the Jews. The greater the economic and cultural handicap, seemingly, the greater the zeal.

(f) One might elaborate indefinitely, but this is sufficient to indicate that a careful study of the individual case is the next step after the rough grouping of a mass intelligence test. It is much easier, and far more satisfying, to deal in broad statistical

trends than it is to tease out the vexatious problems of individual causes in variation. —— But, the latter is the next logical step in testing.

11. Errors in grading undoubtedly play a part in the rank-order differences between the Thorndike rating and the instructors rating. Impressionistic methods of grading, marked bias in weighting certain qualities in work and too few facts for estimate often cause the instructor's grade to be a long way from the absolute facts.

## CONCLUSIONS

1. While there is a fairly high correlation between academic grades and the Thorndike tests, upon analysis the approximation is found to be too rough to permit exclusion, inclusion, or the assumption that the highest test group (above 95) is peculiarly superior.

2. Nevertheless, the college entrance tests are not without merit, offering as they do, an initial point for analysis and judicious prognosis, but unfortunately, not scientific prophecy. Such tests are undoubtedly superior to any impressionistic estimates of the Freshman. It becomes necessary, however, to follow up the test findings with a careful analysis of the qualitative character of the score; to supplement with a common-sense investigation of the nature of the high school training noting how well it dovetails into the curriculum of the particular student in question. The home, language, zeal, outside employment, and so on must be weighed and evaluated. The tests form but a part of a complex differential diagnosis. All factors must be carefully summated analogous to a court's getting a perspective on circumstantial evidence.

3. Whether the student is surpassing or falling below the Thorndike rank-order, he is equally of interest. At present such cases demand an investigation to "show cause."

4. Not only is the Freshman up for critical study and evaluation but the test purporting to evaluate him is equally on trial to show it has value.

5. A combination of Part I, form 5, and Part III, form D would probably be as valuable as all four forms; thus enacting a saving of 50 per cent in giving, scoring, and purchase price of blanks. But it should be added that for purposes of analysis, much is to be gleaned by a study of the qualitative character of all four forms.

6. The rank-order comparison between the Thorndike tests and the academic grades should be continued throughout the college course. At present any assumption as to the superior prophetic value of the mass intelligence test over the Freshman's academic performance, is speculative and idle.

# THE UNRELIABILITY OF M. A. AND I. Q. BASED ON GROUP TESTS OF GENERAL MENTAL ABILITY.[1]

By ARTHUR I. GATES
Teachers College, Columbia University

The practice of computing mental ages and Intelligence Quotients from scores of group tests of general mental ability has been questioned recently on theoretical grounds,[2] and various workers are discovering the unreliability of these measures by comparing the results obtained from different tests.[3] This paper presents data which show that at least under certain conditions, quite ridiculous IQ's may be derived from certain group tests.

The following tests were given on successive days to the pupils of Grades I to VIII inclusive of the Scarborough School: Dearborn, Examination I or II, Haggerty, Delta I or II, Holleys Picture Completion or Sentence Completion, Illinois (III to VIII only), Otts Primary or Advanced, National Intelligence, Forms A and B, (III to VIII only) and Meyers Mental Measure. The Stanford-Binet was given during the year to Grades I to VI inclusive.

The tests were given to groups of twenty or less and particular attention was given to time, instructions, provision with pencils, etc. The pupils had learned to take tests earnestly and we are confident that the test conditions in general were more uniform and satisfactory than will be obtained in ordinary school practice. The results were scored by an expert.

The mental age equivalents of the scores on the several tests were computed by use of the age norm tables accompanying each test.[4] In most cases the line describing the median age score was extended to cover such of our scores as exceeded the highest age norm given. From these data, IQ's were computed.

Table I, which gives the data in detail for Grade I, illustrates the treatment of results. The horizontal lines give the facts for the several subjects; the vertical columns give the facts for the class. Columns 1 to 6 give the IQ's which are summarized at the foot in

[1]This study is a contribution of the Department of Educational Research, Scarborough School, Scarborough, New York.
[2]F. N. Freeman. The Interpretation and Application of the Intelligence Quotient. Journal Educational Psychology. January, 1921. L. M. Terman. Mental Growth and IQ. Journal Educational Psychology. October, 1921, especially pp. 398-401.
[3]For example, J. L. Stenquist, Journal of Educational Research, Dec., 1921.
[4]For the N. I. T., Termans Vallejo norms, published in the Journal of Educational Research, February, 1921, were used.

the form, (1) the highest IQ in the class (2) the lowest, (3) the mean and (4) the mean deviation from the mean. Column 7 gives the mean IQ for each subject; column 8 gives the mean deviation of the subject's various IQ's from his mean and column 9 gives the spread i. e., the subject's highest IQ minus his lowest IQ. These data are summarized at the foot of each column.

The facts of Table I may be considered with reference to certain important school problems.

1. Is the class, on the whole, composed of bright or average pupils? A very general answer — bright — may be given, but the degree of brightness varies widely according to different tests. The mean IQ's range from 109 (Dearborn Examination) to 129 (Myers). According to the one the group should do slightly better than average work, according to the other, very much better. According to the one the average pupil could scarcely complete a college course, according to the other he could do so with ease. Assuming the average chronological age to be six years according to one test the mental age would be 6.5 years corresponding to first grade ability, but according to the other test the mental age would be 7.8 years corresponding to advanced Grade II work.

2. Are the pupils of uniform or diverse brightness? The figures for the lowest IQ are fairly uniform, ranging from 92 to 99 but those for the highest are diverse, ranging from 121 to 165. The mean deviation of IQ's from the average are for the Dearborn 6.0 and for the Myers 12.3, with figures for the other tests scattering between these extremes. Recommendations concerning sectioning for purposes of promotion would depend very greatly on what test is used.

3. What is the IQ of a particular pupil? Pupil A is given IQ's ranging from 104 to 165, that is, by one test he would be classified as "average" by another a "genius." Pupil A has various IQ's ranging from 104 to 144. Column 9 gives the difference in points of IQ between the highest and lowest. The largest difference is 61, the smallest 15, the mean 32. On the average a pupil ranges from "average" to "very superior" intelligence depending on the test. If we take as measures of variability the mean deviation from the average IQ instead of the whole range the figures given in column 9 are of course smaller the greatest being 17.0 the smallest 3.8, the mean 9.3. That is if the average IQ is 118 in the average case half of the IQ's obtained from the several tests will fall between the limits 108.7 and 127.3. Further statistical treatment is unnecessary. It is obvious that the variability of the IQ's obtained by several tests all given within a few days is so great as to make prediction concerning optimum rate

## TABLE I

SHOWING I .Q.'S ORTAINED BY PUPILS OF GRADE I IN SEVERAL TESTS

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| | Stanford Binet | Dearborn Exam. I. | Otis Prim. | Kingsbury | Haggerty Delta I. | Myers M. M. | Mean IQ | Mean Dev. from Mean | Range |
| A | 124. | 104 | 135 | 124 | 125 | 144 | 126 | 9.0 | 40 |
| C | 120.5 | 121 | 144 | 114 | 125 | 131 | 126 | 7.7 | 30 |
| D | 114. | 107 | 122 | 134 | 102 | 122 | 117 | 9.2 | 32 |
| E | 135. | 115 | 138 | 126 | 108 | 121 | 124 | 9.2 | 30 |
| F | 110. | 117 | 119 | 111 | 112 | 129 | 116 | 5.3 | 19 |
| G | 133. | 112 | 132 | 116 | 112 | 125 | 122 | 8.3 | 21 |
| H | 92.5 | 99 | 119 | 105 | 120 | 115 | 109 | 9.5 | 28 |
| I | 104. | 107 | 140 | 116 | 111 | 135 | 119 | 12.5 | 36 |
| J | 96. | 99 | 117 | 104 | 107 | 152 | 113 | 14.8 | 56 |
| K | 103. | 92 | 98 | 98 | 107 | 99 | 100 | 3.8 | 15 |
| L | 120. | 115 | 131 | 127 | 114 | 136 | 124 | 7.5 | 22 |
| M | 108. | 108 | 125 | 100 | 99 | 123 | 111 | 9.2 | 26 |
| N | 110. | 111 | 135 | 118 | 125 | 114 | 119 | 7.5 | 25 |
| O | 104. | 109 | 139 | 117 | 106 | 165 | 123 | 17.0 | 61 |
| P | 121. | 115 | 137 | 98 | 117.5 | 117.5 | 118 | 7.7 | 39 |
| Highest | 135. | 121 | 144 | 134 | 125 | 165 | 126 | 17.0 | 61 |
| Lowest | 92.5 | 92 | 98 | 98 | 99 | 99 | 100 | 3.8 | 15 |
| Mean | 113. | 109 | 129 | 114 | 113 | 129 | 118 | 9.2 | 32 |
| M. D. | 10.1 | 6.0 | 7.6 | 8.9 | 6.8 | 12.3 | 5.4 | 2.2 | 9.6 |

SUMMARY OF RESULTS FOR GRADE II n=19.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Highest | 141 | 147 | 148 | 155 | 150 | 196 | 150 | 17.5 | 76 |
| Lowest | 103 | 97 | 122 | 108 | 92 | 105 | 114 | 5.3 | 21 |
| Mean | 117 | 116 | 132 | 135 | 124 | 135 | 127 | 10.0 | 35 |
| M. D. | 7.4 | 7.3 | 6.9 | 12.4 | 12.6 | 15.4 | 7.6 | 2.8 | 10 |

of promotion, possibility of completing college, etc. quite out of the question.

## RESULTS FOR GRADE II

The results for Grade II are summarized in the lowest section of Table I. For this grade, the group tests, all of them non-verbal, are the same as those used in Grade I.

The results are similar. The IQ's yielded by the Myers test show an average deviation from the mean which is more than twice that given by the Otis. One individual receives an IQ on one test which is 76 points higher than that given by another test; the difference between an average intelligence and idiocy. The average range of IQ for an individual is 35 points.

## RESULTS FOR GRADES III to VIII

Table II gives summaries of the data for grades III to VIII. This Table shows the ridiculous results which teachers and other workers are likely to obtain from efforts to derive M. A.'s and IQ's from group tests in these grades. For example, one fairly bright boy in Grade VII earns a mental age of 46 years on one test giving an IQ of 408. If they should take the results of certain tests seriously there are many pupils who could think rather pitifully of Sir Francis Galton whose IQ is estimated by Terman to be only about 200. According to the Myers test the average IQ's of our Grades VII and VIII is 261.[5] Picture the perplexity of the teacher test enthusiast who finds that pupil K in Grade V earns an IQ of 99 on the National test and 366 on the Myers. Many others are nearly as baffling.

The variability of scores obtained by a given individual becomes greater as the grade becomes higher, as may be seen in Table III which gives the range of IQ's and M. D.'s for the average individual in each grade.

These facts should not be interpreted as a sweeping criticism of the usefulness and validity of group tests of general mental ability. What is urged is a restraint of certain inappropriate uses of these instruments. We should not, it is clear, compute M. A.'s and IQ's from raw scores, but these raw scores themselves may be put to profitable use. It has been shown elsewhere that these group tests at least those very verbal in content give substantial correlations with achievement in school work.[6]

[5]IQ's could not be computed for National and Otis scores for most of our Grade VII and VIII pupils because they earned scores considerably higher than those represented by the line when it becomes horizontal.

[6]A. I. Gates, Correlations of Intelligence tests with achievement in school subjects. Journal of Educational Psychology, March, April and May, 1922.

## TABLE II

SHOWING HIGHEST, LOWEST, MEAN AND M. D. OF I. Q.'S FOR THE GRADE ACCORDING TO THE SEVERAL TESTS. IN LINES 1 AND 2, DIFFERENT INDIVIDUALS MAY BE REPRESENTED BY EACH FIGURE.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Stanford | N. I. T. | Dearborn | Otis | Haggerty | Myers | Illinois | Holley Vocab. | Mean of 7 I. Q.'s | M. D. from (9) | Range (Highest Minus Lowest |
| **GRADE III** | | | | | | | | | | | |
| Highest IQ. | 133 | 127 | 133 | 135 | 147 | 171 | 154 | 166 | 143 | 16.8 | 76 |
| Lowest IQ. | 95 | 94 | 90 | 106 | 100 | 93 | 80 | 90 | 98 | 4.6 | 16 |
| Mean IQ. | 112 | 112 | 109 | 120 | 120 | 120 | 116 | 115 | 116 | 8.8 | 35 |
| M. D. from Mean | 8.9 | 7.2 | 8.3 | 5.9 | 10.2 | 20.0 | 15.3 | 16.2 | 8.7 | 2.2 | 11 |
| **GRADE IV** | | | | | | | | | | | |
| Highest | 140 | 133 | 142 | 399 | 166 | 176 | 162 | 174 | 154 | 61.0 | 292 |
| Lowest | 99 | 100 | 104 | 93 | 113 | 104 | 100 | 104 | 113 | 6.0 | 15 |
| Mean | 120 | 113 | 124 | 168 | 135 | 128 | 126 | 136 | 132 | 17.2 | 73 |
| M. D. | 7.1 | 6.1 | 5.5 | 61.4 | 12.4 | 11.0 | 10.6 | 14.0 | 10.8 | 10.2 | 52 |
| **GRADE V** | | | | | | | | | | | |
| Highest | 139 | 131 | 136 | 169 | 151 | 324 | 156 | 147 | 154 | 55.4 | 266 |
| Lowest | 100 | 100 | 98 | 115 | 113 | 81 | 109 | 112 | 110 | 5.1 | 28 |
| Mean | 116 | 113 | 120 | 139 | 134 | 169 | 132 | 133 | 132 | 19.4 | 88 |
| M. D. | 9.5 | 7.2 | 10.0 | 13.5 | 9.2 | 79 | 13.9 | 7.3 | 12.3 | 11.7 | 59 |
| **GRADE VI** | | | | | | | | | | | |
| Highest | 160 | 141 | 170 | 162 | 164 | 351 | 191 | 152 | 180 | 49.3 | 218 |
| Lowest | 83 | 88 | 80 | 92 | 95 | 70 | 92 | 97 | 90 | 4.8 | 14 |
| Mean | 120 | 115 | 123 | 121 | 130 | 179 | 130 | 125 | 132 | 19.3 | 82 |
| M. D. | 18.8 | 12.9 | 20.2 | 15.9 | 15.5 | 75.6 | 20.6 | 13.0 | 21.8 | 11.9 | 55 |
| **GRADE VII** | | | | | | | | | | | |
| Highest | Not | * | 183 | * | 155 | 408 | 181 | 147 | 210 | 79.0 | 261 |
| Lowest | Given | | 109 | | 110 | 74 | 94 | 104 | 103 | 7.8 | 26 |
| Mean | | | 137 | | 132 | 261 | 138 | 127 | 159 | 43.9 | 146 |
| M. D. | | | 14.2 | | 18.0 | 96.2 | 17.3 | 9.2 | 19.6 | 24.4 | 77 |
| **GRADE VIII** | | | | | | | | | | | |
| Highest | Not | | 148 | * | 138 | 404 | 168 | 141 | 186 | 86.8 | 286 |
| Lowest | Given | | 105 | | 91 | 88 | 85 | 96 | 172 | 5.8 | 20 |
| Mean | | | 124 | | 116 | 261 | 124 | 111 | 147 | 46.9 | 157 |
| M. D. | | | 9.0 | | 8.2 | 56.1 | 14.2 | 8.9 | 13.7 | 16.9 | 54 |

*Scores for nearly all pupils above the equivalent of highest ages given.

TABLE III

| | Range of IQ in the Several Tests for the Average Individual | Mean Deviation from Mean IQ for the Average Individual |
|---|---|---|
| Grade I | 32 | 9.2 |
| II | 35 | 10.0 |
| III | 35 | 8.8 |
| IV | 73 | 17.2 |
| V | 88 | 19.4 |
| VI | 82 | 19.3 |
| VI | 146 | 43.9 |
| VIII | 157 | 46.9 |

## THE CAUSES OF THE UNRELIABILITY OF IQ'S BASED ON GROUP TESTS

An effort has been made to ascertain the causes of the extreme variations in IQ given to the same individual by different tests although we probably have been unable to discover them all or to evaluate accurately those that seem likely.

First, the distribution of native mental ability in the groups upon which the tests were standardized has probably been unequal. That this is often the case was clearly illustrated in constructing norms for the National Intelligence tests.[7] Such differences are not necessarily eliminated by mere accumulation of vast numbers of measures; a fact that is now quite generally recognized. It is probably an important source of inequality of the test norms.

Secondly, the character of school training, especially in reading, writing and arithmetic, affects the results considerably. For example, if certain grades were backward in arithmetic in comparison with pupils upon whom norms were established, they would obtain lower IQ's in those tests which contained relatively large amounts of arithmetical functions other things being equal.

Again results will differ depending on the relative amounts of non-verbal material which the particular tests contain. Two investigations have shown that in Grades III to VIII the correlations between verbal and non-verbal tests are low; sometimes negative in grade groups.[8] In our results the association between

[7] See L. M. Terman and E. D. Whitmire. Age and Grade norms for the National Intelligence Tests. Journal of Educational Research. February, 1921, and G. M. Whipple, The National Intelligence Tests. Journal of Educational Research, June, 1921.

[8] J. P. Herring. Verbal and abstract elements in intelligence examinations. Journal Educational Psychology. December, 1921, and A. I. Gates, same journal, March, April and May, 1922.

verbal and non-verbal tests becomes less close as the subjects become older. This fact is a partial explanation of the fact that the disagreement between IQ's is greater in the upper grades.

The factor of greatest importance in our results is practice in group tests. Table II shows that Scarborough pupils earned IQ's that are obviously too high. On the whole, the more practice preceding a particular test, the higher the IQ's obtained. There is no doubt that practice results in very rapid improvement in such tests.[9] There are three ways in which practice will cause variations in IQ from test to test: First, scores will advance considerably by virtue of general adaptation to test conditions; secondly, the gain will be relatively great in the particular tests which contain large amounts of identical or equivalent exercises; thirdly, the norms for the tests will be high or low, depending on previous practice in intelligence tests as well as on general mental abilities of the subjects on whom the norms were based. Do those who construct norms for their tests actually ascertain the amount, if any, of previous experience in tests their subjects have had? Without a doubt norms established on uninitiated pupils will differ considerably, native ability being equal, from those secured from pupils who have been tested extensively.

Again, the cumulative procedure of establishing norms probably tends in the long run to produce norms that are too low. Testers probably succeed in controlling the time better than other conditions. In the long run numbers of students fail entirely to understand methods of procedure, skip pages, lose time through distraction, broken pencils, etc., etc., whereas the factors conspiring for higher scores are fewer. Scarborough pupils were handled in small groups. They were provided with several pencils. They had learned to do their best in all "tests." Disturbances and misunderstandings very seldom occured. These factors would tend to produce differential success depending on whether a particular test contained more numerous speed or power exercises. Well practiced pupils would probably have a considerable advantage over the inexperienced particularly in the former, because they make a quicker start and maintain a steadier performance.

Finally, most of the Scarborough pupils are of superior intelligence as evidenced by the Stanford-Binet IQ's (which average about 117[10]), by achievement in school subjects and by informa-

[9]For example, see Knight Dunlap and Agnes Synder. Practice Effects in Intelligence Tests. Journal Experimental Psychology, 1920, p. 357-377.

[10]This statement is made because most of the Stanford tests were given before the experience in group tests and because, on retesting this year, it is found that practice in group and other tests has not caused any marked increase in IQ.

tion concerning the social status of parents. It is possible that very bright children with a little practice may exceed average children in these tests by a larger amount than the IQ ratios would suggest. In support of this view we find that some of the very bright pupils very greatly exceed the norms for the grades to which their mental ages correspond in speed of reading. In depth or power of comprehension (Thorndike-McCall) however, their excess over the average, while marked, is not nearly so great. This would make for differential success depending on the relative amount to which achievement in each of the several tests depends upon these abilities.

It will not be profitable to speculate further; what we need is experimentation. These experiences have been sufficient to make one very cautious in the interpretation of group test scores and particularly cautious in the use of the M. A. or I. Q.

The group tests are chiefly useful for purposes of rough classification and for very rough prediction but either should be undertaken only after inquiry into several factors. First, has this pupil ever taken such tests before? If so, which, when, and how many? We need data badly to enable us to make a rough correction in such circumstances. Secondly, what has been the subject's training in school subjects? It was found, for example, that some of our Grade III pupils made enormous improvement in most verbal group tests as the result of having improved in reading during the interval between two tests. We need information concerning the transfer of improvement in all school subjects to achievement in group tests. Finally did the subject lose any time for any reason during the test? This matter assumes great importance in some of the group tests. If all of the pupils have had no (or the same amount of) previous practice in group intelligence tests if their school training has been the same and if no accidents occur the results, at least those of verbal tests, are very useful but coarse instruments for classification.

Perhaps a change from the term "intelligence tests" to some such term as "classification tests" would have a favorable effect in reducing misunderstanding by teachers and laymen and in curbing misuse by over-zealous testers. It has been well said that the work of measurement "may soon need protection from over-extravagant hopes more than from hostile criticism."[11]

[11]E. L. Thorndike. Measurement in Education. The Twenty-first Yearbook of the National Society Study of Education. Part I, p. 8. 1922.

## NEWS AND COMMENT

The Bureau of Vocational Guidance of Harvard University has taken over the publication of the National Vocational Guidance Association Bulletin. It is their purpose to enlarge and improve this Bulletin as support becomes assured for it.

The Journal extends a cordial welcome to this new magazine in the field of vocational guidance. Professor Frederick J. Allen will serve as Editor.

---

## A NOTE ON POPULAR PSEUDO-PSYCHOLOGICAL BELIEFS

Kornhauser and Jackson present data in a recent issue of this journal[1] showing that actual business men do not use quack systems of judging character in their employment work. The results obtained by the use of their questionnaire are rather surprising in view of the apparently successful exploitation of "applied psychology" by various charlatans who invade our larger cities, advertise extensively in the newspapers and on the bill boards, give free lectures in the best hotels or theatres and then conduct large classes for four or six weeks at so much per head. Dr. Joseph Jastrow at Wisconsin has made a special study of these quacks and the writer believes as the result of conversations with Dr. Jastrow plus his own observations in the quack-infected Northwest that the merits of psychological research and applications are imperiled by this wide spread exploitation of pseudo-psychology.

Pseudo-phychological beliefs amounting in some cases to superstitions seem to be more common than most of us realize. The writer in conducting a summer school course in Applied Psychology was amazed to discover the range of interests expressed by these adult summer school students, but even more so, by the beliefs shown in answer to a questionnaire. The range of interests included a desire to study the problem of learning, memory, how to study, business psychology, salesmanship, influencing people, personnel work, personality, vocational guidance, abnormal psychology, psycho-analysis, spiritualism and character analysis. Below is given a summary of the beliefs expressed by these students at the beginning of the course in answer to a questionnaire.

The above table speaks for itself. If mature summer school students exhibit such a variety of beliefs what must the average business man believe? It goes without saying that such a course

[1] "A note for the extent to which systems of character analysis are used in the Business World." Jour. Applied Psy., Vol. VI, No. 3, Sept., 1922, p. 302.

SUMMARY OF BELIEFS EXPRESSED BY APPLIED PSYCHOLOGY STUDENTS

| Do you believe in: | Positive Belief | Positive Disbelief | Unknown or Uncertain |
|---|---|---|---|
| 1. Astrology | 2 | 11 | 7 |
| 2. Chiropractic | 10 | 9 | 1 |
| 3. Fortune Telling | 2 | 18 | 0 |
| 4. Graphology | 5 | 5 | 10 |
| 5. Hypnotism | 12 | 8 | 0 |
| 6. Memory Systems | 12 | 7 | 1 |
| 7. Mental Telepathy | 12 | 8 | 0 |
| 8. Spiritualism | 5 | 15 | 0 |
| 9. Palmistry | 1 | 18 | 1 |
| 10. Phrenology | 5 | 14 | 1 |
| 11. Physiognomy | 11 | 3 | 6 |
| 12. Psycho-therapy | 9 | 1 | 10 |
| 13. Absent Treatment | 1 | 12 | 7 |
| 14. Gifted Children are "Peculiar" | 12 | 8 | 0 |
| 15. Psychology Can Determine What Specific Job a Man Is Best Fitted for | 16 | 4 | 0 |

needs devote some time and attention to evaluating the evidence for and against a large number of pseudo-psychological systems following this up with a presentation of what applied psychology can do and can not do. Thus it turns out that phychological instruction must serve as a sobering survey of the limitations of psychological applications. It may be that business men are free from such false beliefs yet the undersigned classifies himself as a "doubting Thomas."

University of Minnesota DONALD G. PATTERSON.

# BOOK REVIEWS

CLAPAREDE, EDOUARD. *Geneva, Switzerland,* Problems and *Methods of Vocational Guidance.* International Labor Office, 1922. Price, 40 cents.

Interest in vocational guidance has been growing so rapidly that the International Labor Office has taken cognizance of it as a subject of international importance. About two years ago the Office appointed the University of Geneva as its agent to collect information regarding the progress of the movement throughout the world, and to prepare news items for regular publication in the International Labor Review. This work has been carried on in increasing volume and with widening usefulness. In the meantime Professor Claparede has been conceiving and directing investigations fundamental to vocational guidance from his laboratory and from the Institute J. J. Rousseau, and he has brought out this brochure as a precipitation of the results of his study.

It is not a book. It is rather a survey that just escapes the severities of outline presentation. Furthermore it does not treat all phases of vocational guidance. Though it recognizes the importance of the social, economic, physiological and educational aspect it treats specifically only the psychological. This it envisages with great clearness. Particularly sane and proportionate is the critical discussion of the difficulties involved in the application of tests in practical vocational situations. In his summarized conclusions Professor Claparede gives what might be regarded as an international message:

"In conclusion it may be well to summarize very briefly the steps which should be taken to place vocational guidance on a firm foundation.

1. Psychological analysis of occupations and their classification according to the abilities required.

2. Determination of corresponding abilities in individuals and, as a preliminary, the testing of tests to determine their diagnostic, and even more their prognostic, value.

3. Investigation of the process of training and educability.

4. Determination of the importance of general intelligence in the various occupations.

5. It should never be forgotten that a forecast of ability will never be a certainty, but only a probability. Statistics should be prepared to determine this probability.

6. Efforts should be made to arrive at a uniform system of measurement so that psychographs or psychological diagrams may be comparable between different countries and between dif-

ferent vocational guidance offices. Grading by percentiles appears to be the most rational and least ambiguous method.

It is in the social interests of the community that every man should be in the position for which he is best fitted. It is also desirable and increasingly in accordance with the spirit of modern democracy that the various social functions should be allotted according to merit and not in virtue of any privilege. The only means of attaining this ideal is by a sound system of vocational guidance. All those who have at heart the harmonious organization of society, both employers' and workers' organizations, should therefore facilitate by all means in their power any investigation having for its object the establishment of industrial psychology upon a firmer foundation."

Every American phychologist who is touching ever so lightly, either in his teaching or in his investigation, upon vocational problems, should ponder on this brochure. He will find posed for him numerous problems for the experimental investigations so sorely needed, and he will find references to works of European investigators that are too little known in America.

Indiana University H. D. KITSON.

PRESCOTT, F. C. *The Poetic Mind.* New York: The Macmillan Company, 1922. Pp. 308, $2.00.

There are two classes of people to whom this book should be of special interest. The first are the psychologists. Professor Prescott has for the object of his study the working of the poet's mind. His evidence is from the most reliable source—namely, from the poets themselves. An English professor, he has collected from far and wide statements of the poets and other creative artists concerning the source and manner of their inspiration; and with a surprising unanimity of opinion he has found that they testify to a power within from which their creative works emanate and over which they have no voluntary control. This mysterious source of power Professor Prescott attempts to identify with the subconscious mind. He seeks to show further, always depending for his chief evidence on the poets themselves, that the same principles which apply to dreams according to Freud apply also to poems, which are emanations of the subconscious mind just as much as are nocturnal dreams. Poems, like dreams, have their origin in thwarted desires. Step by step Professor Prescott parallels the two kinds of dreams, showing that the impulse, the censorship, displacements, condensations, fusions, and fictional gratification are common to both and have the same explanations. Whatever be the merits of his findings, he has decidedly done the psychologists a service by bringing them into contact with a field of investi-

gation which few of them are likely to be so well qualified as he to explore—the mental operations during poetic inspiration as explained by the poets themselves.

The book, however, should perhaps have even a greater appeal to the critic and the student of literature. In the first place, even if there is not a particle of truth in his thesis, it is a fascinating romance of the subterranean caverns of the human mind. In the second place, if his thesis is partly true, his work becomes valuable and suggestive. It sheds new light on such ancient themes as the relation of beauty to truth, the catharsis of tragedy, the universality of the greatest poetry, symbols and figures, and even puns. In the third place, it is very possible that his thesis is true. Once grant the existence of a subconscious mind of the kind accepted by many modern phychologists and the rest easily follows. The subconscious mind is old and wise, but not accessible to the average individual. The poet is one who for comparatively short intervals does have access to it, though through no volition of his own. What have been regarded as his vague and misty dreams are in reality flashes of truth from a deeper mind, which is racial, universal, unhampered by the limitations of time and space and immediate needs. Poetry, therefore, represents the highest wisdom of the race. Scientists plod along, and by the aid of the conscious mind finally confirm what the poet saw in his vision. Shelley, without the aid of modern psychology, states the idea exactly in his *Defense of Poetry*: "In the infancy of the world, neither poets themselves nor their auditors are fully aware of the excellency of poetry: for it acts in a divine and unapprehended manner, beyond and above consciousness, and it is reserved for future generations to contemplate and measure the mighty cause and effect in all the strength and splendour of their union."

It seems to me, however, that there is at least one rather serious inconsistency in the book. In Chaper VIII Professor Prescott is inclined to accept the conclusion of Wordsworth that emotions recollected in tranquility are "generally" the source of poetry. In Chapter XV he attempts to show that poetry is the result of thwarted desires: that, as the poet may not have the object of his desire, he seeks relief, even though it be a veiled relief, in a fictional gratification, as Keats, let us say, in *The Eve of Saint Agnes*. Such a relief is obviously needed and obviously would come when the emotion is powerful and not when it is recollected in tranquility. Likewise in the last chapter Professor Prescott has no basis for even suggesting the possibility of poets for seeing specific events in a literal sense, as Shelley foresaw his own death by drowning at sea.

Ohio University. B. L. JEFFERSON.

TERRY, PAUL WASHINGTON. *How Numerals Are Read.* Chicago, The University of Chicago, Supplementary Educational Monographs, No. 18, 1922. Pp. XIV+109. Price, $1.00.

Part I of Doctor Terry's excellent monograph deals with studies of the reading of numerals by introspective methods. Part II contains an account of the use of photographic apparatus in dealing with this problem.

It is the opinion of the reviewer that the chief value of the monograph lies in the suggested method of studying the problem rather than in the results of the experiments, since a very limited number of subjects participated. Part II, involving the use of photographic apparatus, contains the records of six subjects; in Part I, the number of subjects ranges in the various experiments from four to ten.

In the first experiment, seven arithmetic problems were solved by ten graduate students who kept notes on their experiences while reading the problems "with special reference to numerals." These observations were organized and recorded in a series of tables. The data made it possible to distinguish between a first reading and a re-reading in many instances as well as a partial and a whole first reading.

Other experiments in which the method of introspective analysis was employed investigated the range of correct recall of numerals after the first reading, undertook an analysis of the re-reading of numerals in arithmetical problems and studied the successes of four subjects in reading numerals in columns.

Important outcomes of the use of photographic apparatus are: "(a) The average number of digits included by a pause on numerals is decidedly smaller than the average number of letters included by a pause on words. (b) The average duration of pauses on numerals is greater than the average duration of pauses on words in the cases of all subjects. (c) The percentage of regressive pauses on numerals is greater than the percentage of such pauses on words."

The final chapter of the monograph contains very pertinent suggestions for teaching children to read arithmetic problems. Inasmuch as the reading of numerals requires closer attention on the part of the reader than does the reading of words, it is suggested that during the first reading pupils be taught to devote their entire attention to the conditions of the problem, developing, at the same time, skill in the partial reading of the numerals. These conclusions are supported by the findings of Gray and Freeman in the field of reading.

Ohio University. R. L. MORTON.

KNIGHT, FREDERICK B. *Qualities Related to Success in Teaching.* Teachers College, Columbia University Contributions to Education, No. 120, New York City. Pp. x+67.

When and why is a teacher a successful teacher? Many attempts have been made to answer this question in terms of such general categories as personality, poise, physical qualities, mental qualities, moral qualities, disciplinary skill, instructional skill, etc. The list can be extended and subdivided almost indefinitely from one of the several score cards for teachers now on the market. The objection that can be made to the above descriptive terms is their generality. Although they are more useful than a "general impression" as a guide to getting a useable criterion of teacher success, more exact and less subjective measures would be welcomed by the school administrator. The study here under discussion represents a contribution to scientific education in that it gives results and particularly an improved method of attack on this problem.

The first chapter summarizes the studies of Meriam, Elliott, and Boyce in this field. Chapters II, III, and IV are concerned with the data of the author's investigation and the relative significance of the various traits and qualities studied. The last chapter gives a trenchant criticism of Boyce's study.

The general procedure of the first part of the investigation was as follows. Ratings as to the relative teaching ability of 156 grade and high school teachers in three different school systems were obtained. That is to say, the best teacher was ranked first, the next best second, etc., when compared with all the other teachers of a given system. Three such rankings on the general ability to teach were secured from the supervisors, the teachers themselves, and a selected group of pupils respectively. By the method of correlation the validity of these rankings, and their relation to age, amount of experience, quality of handwriting, intelligence as measured by test, major academic interests, normal school scholarship, amount of professional study during active service, and ability to pass a professional test ("A Trade Test for Elementary School Teachers" by Knight and Franzen) was established.

A few of the more significant conclusions may be given.

"By using the coefficient of partial correlation we find, in the case of elementary school teachers, that, the factors of intelligence and normal school scholarship being constant, there is a mutual relationship of +.57 between ability to teach and ability to pass a professional test.... The amount of professional study accomplished during active service is also indicative of success in teaching.

"In the case of high school teachers, intellectual differences, as measured by mental tests, appear to be significant. For the selection of high school teachers the use of mental tests would be of value.

"These data, as a whole, may be interpreted to mean that the general factor of interest in one's work becomes the dominant factor in determining one's success in teaching....other measurable traits, either alone or in combinations, are not adequate explanations of teaching success.

"This study indicates that age, experience, quality of handwriting, intelligence as measured by tests, normal school standing, or the expressed interests of teachers are not closely related to success in teaching."

In the second part of the study intercorrelations were found among the traits teaching ability, ability to understand and manage people, intellectual ability, ability to manage things and mechanisms, general scholarship, skill in government or discipline, instructional skill, initiative, nervous and physical strength, ability to command the respect of pupils and general ability to get results, the teachers rating each other on these qualities after the connotations of these terms had been explained to them. Concerning these data the author concludes that the factor of general spread of estimate operates to invalidate such ratings. That is, "when a judge attempts to rate a teacher in some particular trait, his rating is a defense of his general estimate of that teacher, as well as a rating of the trait under consideration."

Doubtless the future will see a scientifically accurate and serviceable measure of teaching ability. The results and methods of such studies as this point the way to the solution of the problem.

Colorado College HERMANN REMMERS.

WALLIN, J. E. W., *The Achievement of Subnormal Children in Standardized Educational Tests.* Miami University Bulletin, Series XX, No. 7. $0.60.

In this bulletin of ninety-seven pages, Dr. Wallin gives the results of some educational tests administered by special class teachers to about 300 St. Louis subnormal children. The tests used were the Ayers spelling scale, lists A to F and I, L and O, and lists I and II in Starch's Spelling scale, Gray's oral reading and the spiral arithmetic excersises adapted from the Courtis tests.

Dr. Wallin states that this "study should have a two-fold significance for educators interested in the study and training of feeble-minded and backward children: first, it is the only survey thus far made by several standardized educational tests of pupils enrolled in special schools for mental defectives; second, most

of the pupils tested had been psychologically examined and differentially diagnosed, so that it is possible to analyze the results from the educational tests with reference to the children's intelligence age, to their diagnostic classification, such as imbeciles, morons, borderline, backward, visual aphasia cases, etc., and to the grade assigned them by the special class teachers."

The results of the educational tests show that it is a waste of time and money to attempt academic instruction to low grade children below four years mentality and that little reading can be accomplished by a child under the mental age of six years. On the other hand the tests show that subnormal children who may develop to about the nine-year mental level should be taken care of in the public schools and that although their training should a reasonable amount of time should be devoted to academic be along the lines of the monumental arts and proper social habits, branches. Dr. Wallin is convinced that it is of extreme importance to take every precaution in placing children in schools and institutions for the mentally defective and that their educational attainment ability must be considered as well as a mental diagnosis made.

Ohio University. M. La Vinia Warner.

*High Schools and Sex Education*: A manual of suggestions on education as related to sex, Edited by Benjamin C. Gruenberg, Ph. D. Prepared under the direction of the Surgeon General, U. S. Public Health Service, in co-operation with the U. S. Bureau of Education. Washington, D. C., Government Printing Office, 1922. Pp. 98. Price $0.50.

This is a very helpful and practical book for teachers of all young people of secondary school age. It makes clear the need of better methods of training in matters of sex. In the very earliest pages it differentiates sex education from sex information—a point which the lay mind seems to have been slow in appreciating. Sex information without interpretation of its meaning to society and the individual and without inspiration to use it for the uplift of the race falls far short of its proper place in education. This little manual has been built up out of the accomulated experience of thousands of teachers who have attempted to handle the delicate subject of sex. It outlines the necessary preparation of teachers of sex education, discusses the type of personality necessary to make such teaching valuable and presents practical methods and devices by which sex education may be accomplished through the regular courses in biology, general science, physiology, physical education, home economics, and the social sci-

ences. Numerous references are cited in order that the reader may have an adequate background for more detailed reading. In the brief appendices at the end of the book are some very valuable suggestions on the use of special lectures, exhibits, courses of study for teachers, personal hygiene and physical examinations. A very comprehensive bibliography of books dealing with sex and sex education ends the book. As a manual of what is being done and can be done in sex education this book deserves a place in every teacher's library.

Ohio University. G. H. GROWDON.

The World Book Company, Yonkers, N. Y., and 2123 Prairie Ave., Chicago, Ill., announces the following:

*1. Otis Self-Administering Tests of Mental Ability*, a new series of tests for grades from four to twelve and for college freshmen. The Higher Examination is for grades nine to twelve and college freshmen, the Intermediate for grades four to eight. Experimental try-outs have been given in preliminary editions. Results have given high correlations with teachers' marks and estimates. Four new advantages, it is argued, are to be had in these tests they are largely self-administering, very easily and quickly scored, afford a flexible time limit, and a wider variety of test material. There are two forms, A and B, each of equal difficulty. The Directions and Key with each package of 25, price $1.00 for Intermediate, $1.10 for Higher Examination.

*2. Pintner-Cunningham Primary Mental Test*: This is a new group test for the kindergarten, first and second grades and the only one on the market suitable for kindergarten pupils. It has been tried out in several previous editions. The advantage of a test applies to the kindergarten and primary grades are that it enables the teacher to classify beginners (with no previous school record), and also to discover by comparison of results, the degree of everlapping in mental ages.

The examination consists entirely of pictures. No reading or writing is necessary on the part of the pupils. The administering is sufficiently simple to require no special training on the part of the teacher.

The examination booklets are sold in packages of twenty-five at $1.45 net per package. The postpaid price of a Specimen Set is 20 cents. One Manual of Directions, one Class Record, one Report to Author and Percentile Graph are included in each package.

*3. National Intelligence Tests, Supplement No. 2.* The Committee on National Intelligence Tests of the National Research Council has prepared Supplement No. 2 to the Manual of Direc-

tions for the tests. This supplement contains revised age norms, based upon the examination of 32,372 pupils in nineteen communities; composition of material for the age norms; data showing variability of the average mental ability of children from city to city; mental age equivalent of scores; revised grade norms, with tables showing composition of the materials used, and a formula for finding IQ's and table for converting years and months into months.

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

BOBBITT, FRANKLIN. *Curriculum-Making in Los Angeles.* Supplementary Educational Monographs, No. 20, June, 1922. University of Chicago, Chicago, Ill. 106 pp.

BONNER, H. R. *Statistics of Teachers' Colleges and Normal Schools.* 1919-1920. Department of the Interior, Bulletin No. 8, 1922. Washington, D. C. 68 pp.

*Consolidation of Schools in Maine and Connecticut.* Department of the Interior, Rural School Leaflet No. 4, 1922. Washington, D. C. 12 pp.

COOK, KATHERINE M. *Supervision of Rural Schools.* Department of the Interior, Bulletin No. 10, 1922. Washington, D. C. 111 pp.

CRAFTS, WILBUR F. *That Boy and Girl of Yours—Sociology from Viewpoint of the Family.* The Baker & Taylor Company, 354 Fourth Avenue, New York. Price $1.75. 432 pp.

DEFFENBAUGH, W. S. *Administration of Schools in the Smaller Cities.* Department of the Interior. Bulletin No. 2, 1922. 75 pp.

DELGADO, HONORIO F. *El Dibujo de los Psicopatas.* Casilla 1589, Lima, Peru. 26 pp.

EDSON, NEWELL W. *Status of Sex Education in High Schools.* Department of the Interior, Bulletin No. 14, 1922. Washington, D. C. 12 pp.

FITT, A. B. *The Human Instincts in Business.* The Lothian Book Publishing Company Pty., Ltd. The Rialto, 497 Collins St., Melbourne, Australia.

FOLMER, ANTH AND GERRITSZ, J. *Tijdschrift Van Den Nederlandschen Werkloosheids-raad.* Bureau: Singel 530, Amsterdam.

GESELL, ARNOLD. *Exceptional Children and Public School Policy.* Yale University Press, New Haven, Conn. 66 pp.

GIVLER, ROBERT CHENAULT. *Psychology—The Science of Human Behavior.* Harper & Brothers, Publishers, New York. Price $3.00. 382 pp.

HEBB, BERTHA Y. *Value of the School Census.* Department of the Interior, City School Leaflet No 3. Washington, D. C.

HOKE, ROY EDWARD. *Improvement of Speed and Accuracy in Typewriting.* Johns Hopkins University Studies in Education, No. 7. Johns Hopkins Press, Baltimore, Md. Price 75 cents. 41 pp.

HOME ECONOMICS IN RURAL SCHOOLS. Department of the Interior. Home Economics Circular No. 13, May, 1922. Washington, D. C. 15 pp.

HOOD, WILLIAM R. *Review of Educational Legislation 1919 and 1920.* Department of the Interior, Bulletin No. 13, 1922. Washington, D. C. 30 pp.

JOHN, WALTON C. *Statistics of Agricultural and Mechanical Colleges for 1919 and 1920.* Department of the Interior, Bulletin No. 27, 1922. Washington, D. C. 29 pp.

*Junior High School Mathematics.* Course of Study and Syllabus. Cleveland Board of Education, Cleveland, Ohio. 45 pp.

KING, LE ROY ALBERT. *Status of the Rural Teacher in Pennsylvania.* Department of the Interior, Bulletin No. 34, 1921. Washington, D. C. 86 pp.

LATHROP, EDITH A. *How Laws Providing for Distribution of State School Funds Affect Consolidation.* Rural School Leaflet, No. 5. August, 1922, Department of the Interior. Washington, D. C. 12 pp.

LAUGHLIN, HARRY HAMILTON. *Eugenical Sterilization in the United States.* Psychopathic Laboratory of the Municipal Court of Chicago. Chicago, Ill. 502 pp.

LOMBARD, ELLEN. *Home Education by Means of Reading Course and the Co-operation of State and National Agencies.* Department of the Interior. Home Education Circular No. 2. Washington, D. C. 11 pp.

*List of References on Rural Life and Culture.* Library Leaflet No. 16, Department of the Interior, Washington, D. C. 10 pp.

PEAR, T. H. *Remembering and Forgetting.* E. P. Dutton & Company, 685 Fifth Avenue, New York. Price $3.00. 242 pp.

*Professional Education of Teachers in Cleveland.* ..Cleveland Board of Education, Cleveland, Ohio. 92 pp.

*Record of Current Education Publications.* Department of the Interior, Bulletin No. 33, 1922. Washington, D. C. 30 pp.

*Reorganization of Home Economics in Secondary Schools.* Department of the Interior, Bulletin No. 5, 1922. Washington, D. C. 38 pp.

*Statistics of Universities, Colleges and Professional Schools 1919-20.* Department of the Interior, Bulletin No. 28, 1922. Washington, D. C. 147 pp.

*Surveys of the Department of Instruction, Cleveland Public Schools.* Cleveland Board of Education, Cleveland, Ohio. 29 pp.

THWING, CHARLES FRANKLIN. *Higher Education in Australia and New Zealand.* Department of the Interior, Bulletin No. 25, 1922. Washington, D. C. 44 pp.

VALDEZAN, HERMILIO AND DELGADO, HONORIO F. *Revista de Psiquiatria y Disciplinas Conexas.* Publication Primestral Vol. IV No. 3. Casilla 1589, Lima, Peru. 262 pp.

WILIAMS, J. HAROLD. *The Twenty-Four Hour School.* Educational Research Bulletin, Vol. I, No. 6, Pasadena City Schools, Pasadena, Calif. 8 pp.

WINDES, EUSTACE E. *A Plan for the Organization of a County System of Agricultural Instruction in Elementary Rural Schools.* Department of the Interior. Rural School Leaflet No. 6. Washington, D. C. 8 pp.

ZOOK, GEORGE F. *National Conference of Junior Colleges, 1920.* Department of the Interior, Bulletin No. 19, 1922. Washington, D. C. 73 pp.

ZOOK, GEORGE F. *Report on the Higher Educational Institutions of Arkansas.* Department of the Interior, Bulletin No. 7, 1922. Washington, D. C. 18 pp.